École Polytechnique Fédérale de Lausanne

COM-480 - Data Visualization

Milestone 2

Students:
Charfi Mohamed
Mrad Ali Ridha
Bouchoucha Mehdi

Professor:
Benzi Kirell

April 25th, 2024

In order to effectively execute our project, it is crucial to divide our objective into smaller, manageable tasks and devise a central visualization highlighting the top-performing players based on various performance metrics. To discern the top players, we must first transform raw statistics into metrics that capture general aspects such as dribbling, defending, physicality, and more. This is how we proceed:

1 Data Preparation

- We begin by first inspecting the datasets and loading them.
- We then handle missing values and preprocess the data by unifying the players' dataset to only have one entry per player (players that switched teams).
- Then we join with the complementary databases to give us additional information for visualization (example: Fifa video game database).
- We then did some web scraping to get the following:
  - Players' images from the website sofifa.com.
  - Club logos from Google Images.
  - Nations flags from Google Images and from a given link in the dataset.

2 Calculating Player Metrics

we created formulas to compute essential metrics like *dribble_metric*, *defensive_metric*, and *aggressive_metric* for each player. Here is an example of how a metric is created:

2.1 Dribbler's Metric

This metric, labeled *dribble_metric*, is a composite measure of dribbling effectiveness for each player in the DataFrame *playerRadarDf*. Here is a breakdown of how it's calculated:

a. Components of the Metric:

- *CarPrgDist*: Total distance, in yards, a player moved the ball while controlling it with their feet towards the opponent's goal.
- *ToSuc*: Number of defenders successfully dribbled past by the player.
- *ScaDrib*: Successful dribbles that lead to a shot attempt.
- *DribblingTotal_fifa*: Dribbling FIFA rate.

b. Calculation:

- Here it is as a mathematical formula: $dribble\_metric = (CarPrgDist + ToSuc + ScaDrib) + 3 \times DribblingTotal\_fifa \times (1 - TeamRank)$

c. Output:

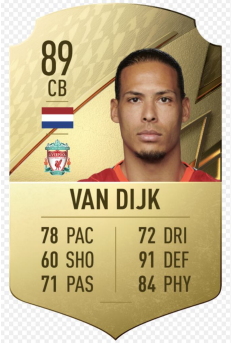
The computed values are stored in the new column *dribble_metric* of the *playerRadarDf* DataFrame. In summary, *dribble_metric* is a weighted sum of a player's carried progressive distance, successful dribbles, FIFA rates and dribble-induced shot attempts, adjusted based on their team's rank. This metric provides a comprehensive view of a player's dribbling effectiveness within the context of their team's performance.

3 Ranking Players

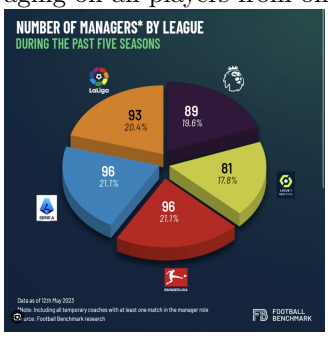
- Combine the calculated metrics to derive an overall grade for each player.
- Sort the players based on these newly calculated metrics to make some visualisations based on those metrics (visualize the best dribblers, the best defenders,etc.)

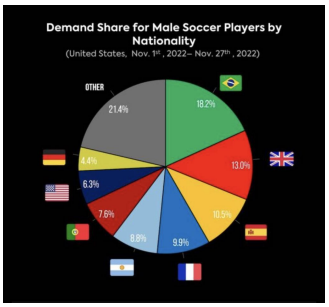
4 Core Visualization

The main ideas that need to be displayed are the following:
Performance score for a player: Standardize raw stats and create new metrics to provide fair scores to each player as the examples below:

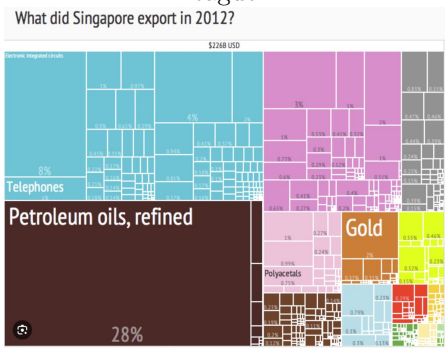


Pie charts: % of dribble success, % of pass accuracy and % of shots target, then averaging on all players from one major football nation or league.





Recursive subdivision chart:
Goals percentage per each team in a league.



Best team of the season:
Assemble the best rated players in a team, compare with journalists' choices.

5 Tools and relevant lectures

The lectures that we will need to complete this project are the lecture 4 to create graphics and visualization with D3.js, lecture 5 to add interactions to our visualisations, lecture 7 and also lecture 11 where we will learn more about tabular data and how to manipulate it. We also have chosen to use Font Awesome in order to create the icons in the website.

6 Additional Ideas for Enhancement

- Implement interactive features like tooltips to provide detailed player statistics upon hovering over data points.
- Include filters or dropdown menus to allow users to compare players from different teams or leagues.
- Develop individual player profiles showcasing their statistics and performance metrics.
- Offer customization options for users to adjust the weightings of different metrics to generate personalized player rankings.