Milestone 1: Project description

"Group U" members : Bakebeck Samuel, Ben Boubker Mahmoud, Jait Fatima-Ezzahra, Mouissa Ismail

Topic : ***Data Visualization in Football Scouting***

Scouting is an essential element in today's football. It allows teams to find the rare gem that can potentially change that team's history.

As a result, teams are ready to invest heavily in this area, even going to create recruitment units in the four corners of the world.

However, a bad judgment on the part of a scout or a bad interpretation of a scouting report by the club board can have catastrophic consequences financially (ranging even around hundreds of millions of euros).

Data

The dataset represents the statistics of the teams of the five big champions (goals scored, conceded, rankings...) and their players (goals, assists, minutes played ...) during the last five years.The teams' dataset come from football-data.co.uk and the players' dataset come from this Github project.Both datasets are in CSV format.

Description of the problem

Users : Scouts, Head of scouting unit, Director of Football

Background and tasks : A scout will supervise players performance and will carry out a report for his head of scouting. The head of scouting will analyze the needs of his team, assign players to be scouted and will forward the best reports to the Director of Football. The Director of Football will have to convince the board to recruit the players. He will also have to negotiate with the player's current team for the transfer fee and the contract with the player's agent.

Purpose of the data : The visualization aimed **primarily** at **exploring the data**. The idea is to analyze and decipher the different methods (data exploration, data processing, data mining and data visualization) to help identify:

- players with similar skills
- potential market value of these players
- position within the teams where these players would be the most effective

The visualization can also be used to justify the purchase of players to a board that lacks football background.

Description of the data

Data size: around 60 megabytes.

Teams Dataset : The tables represent the result of a match between the home and away teams. The dataset has numerical data (goals, assists, shots ...) categorical data (name of the stadium, team, referee) and temporal data (date of matches, season).

Players Dataset : The 1st table represents the performance during a specific season by a player. The data is mainly numerical with the exception of two categorical variables (position played and team). The 2nd table represents the performance by a player during a specific match.