

Football Data Analysis

Group Members:

111903094 Sejal Sarnaik
111903095 Shalaka Pawar
111903113 Trupti Barsagade

Introduction

Football is a widely known sport. From local tournaments to world cups, billions watch and play the game around the world which has produced a lot of entertainment as well as business opportunities to the world. Data Analytics has assumed a huge role in the world of Football. It has transformed how people approach games, team formation, player selection etc. Football analysis was first proposed by Charles Reep in the late nineties to increase popularity of the game.

Data analytics has enabled teams from around the world to understand their game better and perform better. Data analytics is also used to predict the outcomes of games enabling people to make educated guesses while betting. FIFA is Federation Internationale de football association. It was founded in 1904 and is now one of the prestigious associations for football. The game features more than 30 official leagues, over 700 clubs, and over 17,000 players.

The dataset includes the information of the players who were a part of FIFA 2021. It is used to derive insights and correlations between various attributes and gather information.

Business Requirement

1. Clubs have high budget for purchasing players but decision to choose the correct player is always problematic. In the movie Moneyball (2011), the managers of a baseball team were faced with the franchise's limited budget for players. In order to solve this problem, they used a sophisticated statistical analysis approach to analyse

individuals' statistics and scout players. Overall, it was the team composed of undervalued talents that won match after match, therefore demonstrating the success of the analytical approach.

2. Using sensors attached to players and the ball, Football teams are working with data analysis agencies to find out to what extent data can inform decision making in football. In particular, Arsenal has been using data to answer critical questions and forecast the future performance of the team.
3. Optical tracking approach is used to analyse the position of players by AI and ML aspect to form appropriate strategies.

Objectives

1. Find the top 10 highest paid players of the dataset.
2. User can analyze which club or national team had the most unexceptional players based on overall rating attribute.
3. An analysis of the players based on the Value and Wages.
4. Evaluate the co-relation between age and overall rating of the players.
5. Determine which clubs provide the highest salaries to players.
6. Statistically build the dream team from the scraped data based on constraints -
 - a. 15 players in a team.
 - b. Maximum budget.
 - c. Only a fixed number of players from a particular team.
 - d. Attaining players based on positions on the field.
7. Visualization of the comparison between two or more players based on their attributes or skills.
8. Predict player's overall value based on the linear regression and comparison with values given in dataset.

Data Required

1. Basic information about players featured in FIFA 2021.
2. Personal data like Nationality, Age, Weight, Salary, contract details etc.
3. Player, Club, Flag Images to be used in visualization.
4. Player Position Data based on the FIFA 2021 game.

5. Attributes include statistics regarding player moves, skills, preferred footwork, position they served, etc.

The data is scraped from <https://sofifa.com/> website using Beautiful Soup Python library.

The scraped data is stored in CSV format.

Attributes

- Player ID
- Name
- Age
- Photo
- Nationality
- Height
- Weight
- Flag
- Overall
- Potential
- Club
- Club Logo
- Value
- Wage
- Preferred Foot
- Sprint Speed
- Agility
- Weak Foot
- Skill Moves
- Position
- Dribbling
- Stamina
- Strength
- Penalties
- Defensive awareness
- Standing Tackle
- Sliding tackle
- GK Kicking
- GK Positioning
- GK Reflexes
- Positions on field
CAM, CB, CDM, CF, CM, ID, LAM, LB, LCB, LCM, LDM, LF, LM, LS, LW, LWB, Preferred Positions, RAM, RB, RCB, RCM, RDM, RF, RM, RS, RW, RWB, ST

Design Model:

Design Model

