

Executive Summary

Data Collection Project: Football Scores and Statistics

- ***Introduction:***

This data collection project aimed to gather real-time football scores, fixtures, and player statistics from the BBC Sport website for analysis. The focus was on the English Premier League and top scorer statistics.

- ***Data Sources:***

The primary data sources were the football scores and fixtures webpage (for match information) and the Premier League top scorers webpage (for player statistics). URLs for these sources are as follows:

Football Scores and Fixtures: <https://www.bbc.com/sport/football/scores-fixtures/>

Top Scorers Statistics: <https://www.bbc.com/sport/football/premier-league/top-scorers>

- ***Data Collection Process:***

Web scraping techniques were employed to extract data from HTML content. Python libraries such as BeautifulSoup and requests facilitated the extraction of relevant information. The data included match details (teams, scores, and timings) and player statistics (goals, assists, etc.).

- ***Potential Uses:***

The collected data can be utilized for various analyses, including:

- Predicting match outcomes based on historical performance.
- Assessing team and player trends over time.
- Understanding top scorer dynamics and team statistics.

- ***Ethical Considerations:***

While web scraping, ethical considerations were prioritized. Requests were made responsibly to avoid overloading the server. Additionally, the project adhered to the terms specified in the website's robots.txt file.

- ***Limitations:***

Limitations include potential changes to the website structure impacting the scraper's functionality and the need for periodic updates to maintain accuracy.

- ***Accomplishments and Future Directions:***

This semester, significant accomplishments include developing a functional web scraper, organizing and adjusting match time data, and visualizing player statistics. Future directions involve enhancing data visualization and exploring machine learning for predictive analysis.

- ***Conclusion:***

This project provides a valuable dataset for football enthusiasts and analysts. The ethical approach to data collection ensures sustainable use, and the insights gained contribute to a deeper understanding of football dynamics.