# Capstone Project Plan – Technical Write Up

## Abstract

The English Premier League (EPL) stands as a pinnacle of football excellence, captivating millions of fans worldwide with its thrilling matches and fierce competition. In this comprehensive technical paper, we embark on an in-depth exploration of Premier League standings and transfers data spanning over three decades. However, we will only be looking over a decade worth of data from 2010/2011 season to the 2022/2023 season. Our analysis encompasses data retrieval, processing, visualization, and interpretation, providing valuable insights into club performances, trends, and the impact of external factors such as the COVID-19 pandemic. Additionally, we elucidate the code implementation behind the analysis, offering transparency and reproducibility for future investigations.

**Key words:** Premier League, Standings Data, Transfer Data, COVID – 19 Impact

## Introduction

The English Premier League (EPL) stands as a beacon of excellence in the world of football, renowned for its intense competition, global appeal, and high-stakes matches. Analyzing Premier League standings data offers a window into the league's dynamics, providing stakeholders with crucial insights into club performances, trends, and the factors shaping success. In this paper, we embark on a comprehensive analysis of EPL standings data, aiming to uncover hidden patterns, identify trends, and shed light on the league's evolution over the past decade.

## Data Retrieval and Processing

To initiate our analysis, we leverage the API-Football database to retrieve historical Premier League standings data for the seasons spanning from 2010/2011 to 2022/2023. The data includes a plethora of metrics such as matches played, wins, draws, losses, goals scored, goals conceded, points accumulated, and additional descriptors for each club in the league. We employ Python scripts to retrieve, process, and store the data in a structured format for subsequent analysis.

## Code Implementation

Below, we provide an overview of the Python scripts utilized for data retrieval, processing, and visualization:

### Data Retrieval

A screenshot of a computer

Description automatically generatedWe utilize the **requests** library to make HTTP requests to the API-Football database, fetching Premier League standings data for each season within the specified timeframe. The API key is stored securely, ensuring authentication during data retrieval.A screenshot of a computer program

Description automatically generated

### Data Processing

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generatedOnce the data is retrieved, we process the JSON response, extracting relevant information such as club standings, points, goals, and descriptors. We construct Pandas Data Frames to organize the data for further analysis and visualization.

# Data Cleaning

Our data cleaning process involved identifying and removing redundant, duplicate, and erroneous data. We meticulously documented each step of the data cleaning process, ensuring transparency and repeatability.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### Data Visualization

A screen shot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generatedTo enhance understanding and interpretation, we employ various visualization techniques such as bar charts, line plots, and comparative analyses. These visualizations offer insights into club performances, trends, and comparisons over time.

A screenshot of a computer code

Description automatically generated

A screenshot of a graph

Description automatically generatedA graph with colorful lines and dots

Description automatically generatedA graph with lines and dots

Description automatically generated

## Key Findings

Our analysis uncovers several key findings, including insights into top clubs' dominance, relegations, COVID-19 impact, and participation in European competitions. Visualizations aid in the interpretation of trends and patterns, providing stakeholders with actionable insights for decision-making.

## Conclusion

In conclusion, our analysis of Premier League standings data offers valuable insights into club performances, trends, and dynamics within the league. By combining data retrieval, processing, visualization, and code implementation, we provide a robust framework for future analyses and investigations in the field of sports analytics.

Future Directions

Future research endeavors could explore additional variables such as player statistics, managerial influences, and tactical trends to enrich our understanding of Premier League dynamics. Furthermore, longitudinal studies tracking club performances over extended periods could unveil deeper insights into long-term trends and patterns.

## Acknowledgements

I express gratitude to the API-Football database for providing access to Premier League standings data, enabling our analysis. Additionally, we acknowledge the Python community for developing powerful libraries and tools that facilitate data retrieval, processing, visualization, and analysis.

## References

<https://api-football-v1.p.rapidapi.com/v3/standings>

<https://api-football-v1.p.rapidapi.com/v3/transfers>

This technical paper serves as a comprehensive exploration of Premier League standings data, offering insights, trends, and methodologies for future investigations in sports analytics. Through transparent code implementation and robust analysis, we contribute to the broader understanding of one of the world's most prestigious football leagues.

This expanded technical paper provides a comprehensive overview of the analysis conducted on Premier League standings data, offering a deeper understanding of club dynamics, trends, and the league's evolution over time. Through meticulous data retrieval, processing, visualization, and interpretation, we shed light on key insights and contribute to the body of knowledge in sports analytics.

## Additional Visualizations

A graph with red bars

Description automatically generated

A chart with numbers and a number of different colored bars

Description automatically generated with medium confidence

A graph of blue bars

Description automatically generated

A graph with numbers and a bar

Description automatically generated

A graph of a match

Description automatically generated with medium confidence