Problem Statement Worksheet (Hypothesis Formation)

- **"Develop a predictive model by May 19, 2024 (end of the season), to forecast a team's league performance for the upcoming season, utilizing historical data on squad statistics (including mean age, number of foreign players) and financial investments (salaries and spending) from the 2015-2016 season through the 2022-2023 season. The model aims to predict the team's end-of-season rank and total points with an accuracy improvement of at least 10% over baseline predictive models. This analysis will provide team's league management with insights into the impact of squad composition and financial strategies on performance outcomes, aiding in strategic planning and decision-making to optimize future team success."**

SMART

- **Specific: The project now aims to develop a predictive model applicable to any team, forecasting league performance based on detailed historical data, making the goal broad yet specific in its requirements.**
- **Measurable: The success criterion remains clear and quantifiable, focusing on achieving at least a 10% improvement in prediction accuracy compared to existing models, alongside predicting end-of-season rank and total points.**
- **Action-Oriented: Given the timeline extended to the end of the season in May 2024 and the availability of historical data from 2015-2016 to 2022-2023, the goal of developing a robust predictive model is realistic with dedicated research and analysis.**
- **Relevant: This project is highly relevant to team management across the league, offering actionable insights that could significantly influence strategic decisions regarding squad composition and financial management.**
- **Time-bound: The project has a specific deadline (May 19, 2024), aligning with the end of the football season, which provides a clear timeframe for achieving the stated objectives.**

Context (Why are you working on this problem?):

- **The project is initiated to harness the predictive power of soccer, a sport where strategic decisions significantly influence success. With the evolving landscape of soccer analytics, understanding the interplay between squad composition, financial investment, and performance outcomes has become crucial. This project seeks to provide a data-driven foundation for these strategic decisions, enhancing competitive edge and optimizing team success through informed planning and resource allocation.**

Criteria for Success (What are the key criteria that will deem this work successful?):

- **Successful development of a predictive model that forecasts a team's league performance for the upcoming season with a specific focus on predicting end-of-season rank and total points.**
- **Achievement of at least a 10% improvement in prediction accuracy over existing baseline predictive models.**
- **Provision of actionable insights to team management on optimizing squad composition and financial strategies for future success.**

Scope of Solution Space (What is the focus of this business initiative? What specific items will you focus on exclusively?):

- **Utilizing historical data from the 2015-2016 season to the 2022-2023 season for analysis.**
- **Analyzing squad statistics such as mean age and the number of foreign players.**
- **Reviewing financial investments, including player salaries and spending on transfers.**
- **Non-quantifiable psychological factors affecting player performance.**
- **Short-term market dynamics that might impact player valuation and salaries.**

Constraints within Solution Space (What constraints exist that may prevent this business initiative from succeeding?):

- **Limited access to comprehensive financial data due to confidentiality and privacy concerns.**
- **Variability in data quality across different sources, which could impact the accuracy of predictive models.**
- **Potential biases in historical data that might not accurately reflect future conditions or unforeseen impacts (e.g., global events).**

Stakeholders to Provide Key Insight (Who are the key stakeholders that need to be involved in this project? Where will you source your data from, and who will you present your recommendation to?):

- **Soccer club management teams for providing insights and verifying data accuracy.**
- **Data scientists and sports analysts for developing and validating the predictive model.**
- **The strategic planning departments of soccer clubs for implementing recommendations.**
- **Data sources will include public databases, sports analytics companies, and possibly proprietary team data.**
- **Recommendations will be presented to the executive boards and coaching staff of soccer clubs.**

Key Data Sources (What are the key pieces of data you need to answer the questions related to the problem you are trying to solve?):

- **Historical performance data of soccer teams, including league standings, points, and goal statistics.**
- **Detailed squad statistics for each season, covering player ages, nationalities, and roles.**
- **Financial records detailing expenditures on player salaries and transfer fees.**
- **External factors influencing soccer leagues globally for context in predictive modeling.**