Datasets we already have:
- Transfermark for player transfers, fees, and valuations.
- Understats for player performance stats like goals, assists, minutes played, etc.

Datasets we are waiting for permission to use:
- Sofascore
- WhoScored?

Both of these would give us more performance stats than understats, which mostly gathers offensive stats. So if we do not get permission to use these data sources, we might need to focus on attacking players in our analysis.

Objective:
Develop a model that predicts the optimal transfer value of a soccer player based on historical performance and financial data, ensuring that clubs make financially sound investments that maximize return on investment and on-field success.

Stakeholders:
Soccer clubs that want to optimize transfer spending. They could use the model to avoid overpaying for players, identifying undervalued talent instead. The model would allow clubs to invest in players whose performance metrics translate into better on-field outcomes, maximizing their ROI. By aligning player acquisitions with financial goals, clubs can better manage budgets.

KPIs:
- On-field ROI: How much return a club gets from a player's performance based on their transfer fee?
- Player valuation accuracy: Measure the difference between the predicted optimal player value from the model and actual market prices.
- Club on-field success: Is the club more successful after the transfer (better placed in the league)?