

**Final Project Report**  
**Optimizing Player Acquisition Strategies for Soccer Clubs**

Hiba Hassan - hibah - [hibah@andrew.cmu.edu](mailto:hibah@andrew.cmu.edu)

Contribution: Question Analysis, Report, and Presentation

Niharika Patil - hiharikp - [niharikp@andrew.cmu.edu](mailto:niharikp@andrew.cmu.edu)

Contribution: Data Cleaning and Modeling, Report, and Presentation

Genna Moellering - gmoeller - [gmoeller@andrew.cmu.edu](mailto:gmoeller@andrew.cmu.edu)

Contribution: OptiGuide, Question Analysis, Report, and Presentation

94867 - Decision Analytics for Business and Policy

Dr. Peter Zhang

December 8, 2023

## Executive Summary:

*Background:* Football is one of the most popular sports in the world, and its competitiveness transcends what the players do in a given match. Coaching staff, football operations, and general managers all do their part to maximize their teams' advantages before the players even take the field. One of the primary ways they can do this is by acquiring the best players they can afford. The conventional process for finding and acquiring the players best suited to a team can be long and arduous, as various inputs like performance statistics, risk assessments, and team need analysis need to be factored in. This environment is ripe for data-driven decision making.

*The Problem:* Our analysis draws inspiration from the problem made famous in *Moneyball*, the book by Michael Lewis that narrates how a Major League Baseball team overcame the competitive gap between rich and poor teams by identifying undervalued players with underappreciated performance metrics that could significantly contribute to their team's success. This report uses data from the European Football Transfer market to conduct a similar analysis for a hypothetical European football club, using an optimization model to help determine which players they should target as trade prospects under given constraints.

*Data, Assumptions, and Modeling:* Our dataset was sourced from TransferMarkt.com, this multidimensional dataset allows for an in-depth exploration of player performance, valuation trends, and strategic considerations across different playing positions in the realm of football. Some key assumptions integrated were budget availability, pool of players to transfer selected from 2023. Emphasizing a holistic approach to team formation our model puts importance on factors such as experience, current form, and overall performance metrics. In terms of our modeling, we first formulated a comprehensive performance metric. Then we opted for a Mixed Integer Problem, where the objective of the model is two-fold: maximizing the value added to the team through player transfers while simultaneously optimizing performance and minimizing the cost of transfers.

*Challenges and Constraints:* Given data constraints and modeling assumptions, our key challenges concerned how data-driven and realistic our model could be. Factors such as lack of transparency from clubs regarding budget and player needs limited integration of qualitative factors. OptiGuide itself had accessibility issues and limitations in responses, particularly when dealing with significant shifts in topics. Acknowledging and addressing these challenges will be key to refining and enhancing the model's capabilities in future applications.

*Conclusions and Future Work:* Overall, we were pleased with OptiGuide and our model's performance. It did fail to meet the demands of several of our questions, but, overall, OptiGuide responded decently to our diverse queries, addressing concerns about excluding specific players, and adjusted budget constraints. We noted how superior GPT-4 was to its GPT 3.5 counterpart in handling our queries, and moving forward, we would like to increase our model's ability to handle prompts that are more vague and respond with important detail like player names and relevant data. In sports, both quantitative performance metrics and qualitative, intangible elements like player chemistry and star power are important. We see OptiGuide as a unique tool capable of bridging the gap between experts in each arena by making the optimization process more accessible with plain language.

## Analytical Model and Data:

*Problem Statement:* As a General Manager of the FC Tartan Soccer club with a budget of 150 million Euros, which athletes should I focus on acquiring given my team's budget, position needs, and the pool of available players?

*Data and Assumptions:* The simplest approach to solving this problem is by using a mixed integer optimization algorithm with Gurobi. We started by looking at the available dataset and conducting exploratory data analysis.

*Data Summary:* Our dataset from TransferMarkt.com, accessed through dataworld (<https://data.world/dcereijo/player-scores>), provides information on players, clubs, competitions, and valuations, which can inform decisions on what players are worth investing in.

*Data Cleaning Summary:* Leveraging the datasets available, we used player demographics, player valuations, and player appearances datasets. The details of the datasets are described below. We started by including only the players active in the last season of 2022 and their most recent valuations. We calculated their performance metrics over different periods. We then got rid of the rows with nulls (about 3% of the data) and converted the columns to relevant datatypes. We also conducted some exploratory analysis to describe the datasets. The focus of the study for this project was limited only to players in the positions 'Attack', 'Defense', and 'Midfield'. Since the performance metrics were available only at a high level, we could not perform a more detailed analysis to include sub-position in our considerations.

*Players:* This dataset contains information about each of the players

- player\_id, current\_club\_id, country\_of\_citizenship, position, foot, height\_in\_cm, market\_value\_in\_eur, highest\_market\_value\_in\_eur, contract\_expiration\_date, current\_club\_domestic\_competition\_id

*Player Valuations:* This dataset contains information about each player's market value and relevant club affiliation

- date, player\_id, current\_club\_id, market\_value\_in\_eur, player\_club\_domestic\_competition\_id

*Appearances:* This dataset contains information about player performance in each of the games the players played for each club they were part of.

- game\_id, player\_id, player\_club\_id, date, player\_name, competition\_id, yellow\_cards, red\_cards, goals, assists, minutes\_played

Due to our limited performance metrics, we had to develop a performance index factoring. The details are described in the next section.

## Analytical Formulations:

The GM of the Tartans soccer club values players' performance in the most recent years, their "form" or the time decay factor, and their overall career experience.

### *Part 1. Developing a performance-scoring metric:*

First, we used a simple weighted average for measuring the cumulative interim performance scoring for players in the playing positions (attack, midfield, and defense). We aggregated the performances for their entire career, years 2021 and 2022, and one for years 2019 and 2020 using the formula below.

Interim Performance score for 'Attack' position for t time period:

$$\text{Performance Score} = (\text{Goals Scored} \times 0.3) + (\text{Assists} \times 0.25) - (\text{Yellow Cards} \times 0.05) - (\text{Red Cards} \times 0.1) + (\text{Minutes Played} \times 0.1) + (\text{Games Played} \times 0.1) + (\text{Minutes per Game} \times 0.1)$$

Interim Performance score for 'Midfield' position for t time period:

$$\text{Performance Score} = (\text{Goals Scored} \times 0.3) + (\text{Assists} \times 0.25) - (\text{Yellow Cards} \times 0.5) - (\text{Red Cards} \times 0.3) + (\text{Minutes Played} \times 0.1) + (\text{Games Played} \times 0.1) + (\text{Minutes per Game} \times 0.1)$$

Interim Performance score for 'Defense' position for t time period:

$$\text{Performance Score} = (\text{Goals Scored} \times 0.1) + (\text{Assists} \times 0.1) - (\text{Yellow Cards} \times 0.25) - (\text{Red Cards} \times 0.3) + (\text{Minutes Played} \times 0.1) + (\text{Games Played} \times 0.1) + (\text{Minutes per Game} \times 0.1)$$

We then use these score to calculate the performance factor and time decay factor of the final metric as follows:

$$\text{Performance Factor} = \text{Interim Performance score for career} * 0.5$$

$$\text{Time Decay Factor} = (\text{Interim Performance score for 2022}) * 1.5$$

$$+ (\text{Interim Performance score for 2020}) * 1$$

The experience factor is calculated as follows:

$$\text{Experience Factor} = (\text{No of years of experience in professional football} / \text{Age in 2022}) * 0.5$$

The final performance score is calculated as the sum of the three factors.

$$\text{Score} = \text{Performance Factor} + \text{Time Decay Factor} + \text{Experience Factor}$$

Finally, we normalized the score on a 0-50 scale to facilitate comparison, and better interpretability.

### *Part 2. Optimization Model:*

In the optimization model, we developed using Gurobi, we seek to identify the most efficient football team that fulfills our requirements within a set budget. It maximizes a combination of the players' performance and value while minimizing the cost of the transfer.

The formulation is as follows:

$x_{ij}$  is a binary variable representing whether player  $i$  is transferred to team  $j$ .

$C_{ij}$  is the cost of transferring player  $i$  to team  $j$ .

$V_i$  is the value of player  $i$

$P_i$  is the performance score of player  $i$  in position  $k$ , as calculated above.

$N_{jk}$  is the need for position  $k$  in team  $j$

$B_j$  is the budget for team  $j$ .

**Objective: Maximize**  $\sum_{i,j,k} (V_i * P_{ik} * x_{ij}) - \sum_{i,j} (C_{ij} * x_{ij})$

**Subject to:**

1. **Budget constraints:**  $\sum_i (C_{ij} * x_{ij}) \leq B_j$  for each team  $j$
2. **One team per player:**  $\sum_j x_{ij} \leq 1$  for each player  $i$
3. **The team needs**  $\sum_i P_{ik} * x_{ij} \geq N_{jk}$  for each team  $j$  and position  $k$ .
4. **Binary constraints:**  $x_{ij} \in \{0,1\}$

### **NOTE:**

1. The model is built to handle identifying players for multiple teams. However, we have tried test case scenarios for one team.
2. The market value  $V_i$  was used as a proxy  $C_{ij}$  as the cost of transfers was not publicly available.

## Analysis with the Language Wrapper:

We asked our model questions that forced it to adjust constraints, change the objective, and explain the reasoning behind its decision-making process. Of note, we ran the notebook several times to check the consistency of our responses, and the answers we got at each trial were similar but not always identical. However, we did get some different responses to our questions when we changed the order in which we asked them or changed our verbiage slightly. In these cases, the optimization portions remained the same, but the language surrounding the answer could vary based on sequencing. Our top 10 questions are outlined below:

### 1) What is our optimal player lineup if we want to trade for at least one attack position, one defender position, and one midfield position and why should we select those players?

#### Selected OptiGuide Response:

*The model optimized player transactions for the desired football team and displayed the lineup of players that would generate the maximum performance, value and stay within the budget limit.*

*The optimal team formation according to the executed code would be the following. Please note that player\_id doesn't refer to actual player's dialogue but they are database identifiers for the players:*

*(Example: Replace this part with actual player data returned from your code execution.)*

*```*

*Team 131*

*player\_id : name : position : normalised\_performance\_score : market\_value\_in\_eur*

*1 : John Doe : Attack : 0.9 : 10000000*

*2 : Jane Doe : Midfield : 0.8 : 20000000*

*3 : Tom Doe : Defender : 0.85 : 15000000*

*```*

*By implementing at least one player for each requested position (Attack, Defender, Midfield), we have adhered to the requirements of having at least one player of each type. The players for each position were selected based on a balance of cost (market value), and performance (normalized performance score).*

*This tradeoff ensures that we obtain the best possible team by maximizing player performance and market value and minimizing the cost, keeping our transactions within the predefined budget.*

*Also, note that the optimal solution has an objective value of 7270951332.977353. This outcome is the same with the original result, which indicates a successful optimization process.*

#### Question Intent and Model Performance:

With this question, we attempted to get a baseline result from our model that could act as a springboard for future questions and/or adjustments. We wanted it to provide a description that included the players in the optimal solution and at least some justification for why those players were selected over the others. The selected OptiGuide response above did well when we put this question at the beginning of our sequence, but asking this question in the middle or toward the end of our 10 questions sequence led to an output with more confused language - although both responses selected the same set of players. One thing we were disappointed with was that the OptiGuide response only created a placeholder for the actual player data returned from code instead of inserting it into its output on its own. For reference, we included the player data below:

	player_id	name	position	normalised_performance_score \
212	36139	Dusan Tadić	Attack	47.619493
1978	38253	Robert Lewandowski	Attack	49.565018
2069	65230	Sergio Busquets	Midfield	48.859354
2099	15452	Raúl Albiol	Defender	45.440993
2123	59561	Dani Parejo	Midfield	49.273872
2484	148455	Mohamed Salah	Attack	50.000000
2972	54781	Nicolás Otamendi	Defender	43.884069

As we can see, the model successfully accommodates our constraint to include at least one player of each position type. The model appears to have chosen players that offer a good balance of market value and overall performance. However, the model is unable to take into account aspects of personal preferences, team dynamics and player health. This, essentially, is one of the model's primary limitations, as such qualitative identifiers are hard to track or numerically inculcate into the model itself.

## **2) We do not like player 15452, who should we hire if we remove him from consideration?**

Selected OptiGuide Response:

*Sure, let's analyze the results.*

*After excluding player 15452 from consideration, the optimization model was re-run. The new optimal objective value, after recalculating with the new constraints, is 7270586358.72 euros.*

*In comparison, the optimal objective value with the original constraints was 7270951332.98 euros.*

Question Intent and Model Performance:

This question was designed to assess our model's ability to take into account intangible factors like team dynamics. The use case for such a question would be if a coach, manager, or other stakeholder saw one of the players returned by the optimization and immediately ruled them out due to an intangible factor like propensity for conflict with existing players, behavioral

problems off the field, the player having previously expressed having no interest playing for the team, etc.

When we first ran this question using the player name instead of the player\_id value, we got an intelligible response from OptiGuide saying that it removed him from consideration. However, when we referenced the actual optimization results, we found that our excluded player was actually still in the line-up. When we tried again using the player\_id number, the selected output above was returned and the optimized results did not include player 15452 (Raúl Albiol). We also found that it was necessary to explicitly preface the number with the word 'player.' When we didn't, there were occasions when OptiGuide would get confused and grab a player that had another numeric value (such as playing time) that matched the player number we inputted, and use that row instead.

### **3) We definitely want player 38253, who else should we hire if he absolutely has to be in the lineup?**

#### Selected OptiGuide Response:

*Incorporating player 38253 into the lineup, the new optimal solution was found with an objective value of 7270951332.977353. This is an improvement from the original solution, which had an objective value of 7209217116.777735.*

*This suggests that including player 38253 not only meets your request but it also enhances the team's overall performance. After including this player, the optimization model will select other players to round out the team lineup and stay within the budget constraints.*

#### Question Intent and Model Performance:

In the converse of the previous question, we wanted to see if we could make sure that a specific player was in the line-up. This might be applicable in a real-life context if a team wanted to build their scouting strategy around a particular player. This question gave us quite a bit of trouble, even though we did successfully get OptiGuide to include Robert Lewandowski in the optimal solution in one of the iterations. When we ran the exact same prompt just in a different place in our questioning sequence, the model either could not find the player and produced an output of “*Sorry I cannot answer your question*” or produced an output saying it had adjusted to include Lewandowski when the optimization output still did not have him in the lineup. We tried several other ways of phrasing the question in which we included Lewandowski's name and player\_id, and just his player\_id, but none of those prompts were able to recreate the first iteration that produced the output we wanted. The phantom output from OptiGuide

We chose Lewandowski specifically because he had the most goals in 2022, and we wanted to see how the model would adjust when forced to incorporate his relatively high market value. When the model did include him in one of our sequencing attempts, the new optimal solution was found to be 7,270,951,332.98. This was an overall improvement from the base solution, reflective of the fact that incorporating Robert enhances the team's overall



performance. The model selected Dusan Tadić, Sergio Busquets, Raúl Albiol, Dani Parejo, Mohamed Salah and Nicolás Otamendi alongside Robert Lewandowski in the optimal solution.

**4) Our budget decreased by 5 million euros, how does that change which players we should try to trade for?**

Selected OptiGuide Response:

*After adjusting the budget by decreasing it by 5 million euros, the optimization model provided a new set of players to trade for a new objective value of approximately 7,047,594,290.77.*

*This objective value is lower than the original objective value (7,270,951,332.98) due to the reduced budget.*

*These results mean that the decrease in budget has led to an adjustment in the team composition, which has a slightly lower overall performance score and cost, meeting the constraints of the new budget. The specific players to be selected for the trade can be found in the `selected\_players` table printed by the code.*

Question Intent and Model Performance:

In this question, we wanted to see what our model would do when we tightened the budget constraint and if that would lead to meaningful changes in our original line-up. There are a number of reasons why a team may need to adjust this number throughout the scouting and trade prospecting process. Being able to adjust this fluidly would allow the club operations teams to rapidly test a variety of different financial scenarios and see how their potential value would change. However, this question gave us some problems as well. Again, the language in the output in our first attempt implied that OptiGuide understood what we were asking but the output in the code didn't line up with its response. In one of our iterations, when we attempted to rephrase and make our question more specific, "Our budget is now 145 million euros, how does that change which players we should try to trade for?", and the output was as follows:

*The budget decrease from €150,000,000 to €145,000,000 has caused the optimization model to revise its player selection strategy. The new objective value is €7047594290.77, compared to the previous objective value of €7270951332.98. This means that the overall performance score times market value has slightly decreased due to the lowered team budget.*

*The new selection of players that our model is suggesting are:  
(Here you would substitute the answer with the actual dataframe output of the last cell of the code snippet)*

*Please review these players and their attributes, considering our budget constraints and team needs. If you have any further inquiries or adjustments to consider, feel free to ask.*

The results in the dataframe provided players whose combined market\_value\_in\_eur added up to 154 million euros and not 145 million like we had specified in our prompt, which

almost looked like the optimization model added 5 million euros to the budget instead of subtracting it.

**5) Our budget increased by 10 million dollars, how does that change which players we should try to trade for?**

Selected OptiGuide Response:

*After increasing the budget by 10 million euros, the team was able to select different players that improve the team's overall performance.*

*The total performance score (objective value) for the team increased from approximately 7.209 billion to approximately 7.719 billion with the additional budget. This indicates that with more budget, the team could afford players with higher performance scores, thereby increasing the overall team's performance.*

*The detailed information about the new team selection is pending. You can review the selection details (which include player\_id, normalized performance score, and market value in euros) for the precise changes in player selections with the increased budget. To get comprehensive understanding of the changes, compare the newly selected players to the originally selected ones.*

Question Intent and Model Performance:

This question was asked in a similar spirit to the previous one, only we increased the budget rather than tightening it. Unlike the previous question, OptiGuide performed well with this one, and selected players with market valuations adding up to 160 million. As expected when more resources are available, the total performance score improves from approximately 7.209 billion to approximately 7.719 billion. As an increase in budget allows selection of players with higher performance scores therefore, an improvement in the overall teams performance. The players selected for this line up are as follows:

	player_id	name	position	normalised_performance_score \
212	36139	Dusan Tadić	Attack	47.619493
1594	57500	César Azpilicueta	Defender	45.207796
1978	38253	Robert Lewandowski	Attack	49.565018
2069	65230	Sergio Busquets	Midfield	48.859354
2071	69751	Jordi Alba	Defender	45.598414
2099	15452	Raúl Albiol	Defender	45.440993
2123	59561	Dani Parejo	Midfield	49.273872
2484	148455	Mohamed Salah	Attack	50.000000

There was one additional member, César Azpilicueta, added to this line-up. Nicolas Otamendi was also dropped from the original lineup in favor of Jordi Alba.

**6) Why should we select player 36139 over player 169880?**

### Selected OptiGuide Response:

*Sure.*

*Looking at the data for the two players, we have the following information:*

*Player 36139:*

- Performance Score: Value 1*
- Market Value: Value 2*

*Player 169880:*

- Performance Score: Value 3*
- Market Value: Value 4*

*We must look at these parameters because the model aims to maximize the normalised performance score while staying within a pre-determined budget (represented by the market value). Therefore, a player with a higher performance score and a lower market value is more beneficial for the team.*

*Based on this, we can derive several possible explanations:*

- 1. If Player 36139 has a higher performance score and his market value is less than or equal to that of Player 169880, this means Player 36139 provides a better "value-for-money" deal. Thus, he is chosen over Player 169880.*
- 2. If Player 36139's performance score is higher than Player 169880's, even if Player 36139 costs more, the model might still prefer him due to the aim of maximizing total team performance.*
- 3. If Player 36139's market value is lower even though his performance score is not higher, the model might select him to allow for other, more costly transfers while not sacrificing too much performance.*

*Lastly, it's important to note that these are simplified explanations. The model's decision is influenced by the entire team's composition and the interplay of many variables and constraints.*

*Comparing the new objective value of 7270951332.977353 with the original objective value of 7209217116.777735, the model has increased the team's performance by effectively selecting the players, including Player 36139.*

### Question Intent and Model Performance:

With this question, we wanted to evaluate how the model would compare two players that appear in one optimization result and not another. This is another feature we considered useful for team operations and analysts considering various transfer combinations. In this case, the model appears to do a sound job providing its reasoning for selecting Dusan Tadić (Player

36139) over Giovanni Di Lorenzo (Player 169880). It does so by highlighting the comparison between their performance scores and overall market values. Given the model constraints, we would prefer a player that has a higher performance score and a lower market value as they would be the most beneficial for the team. The model helpfully poses three scenarios under which we may select Dusan over Giovanni. In the first case, Dusan offers a higher 'value for money deal' where he has a higher performance score and his market value is less than or equal to Giovanni.

In the second case, Dusan is more expensive than Giovanni, but his performance score is the greater of the two. Here, the model would select Dusan. In the final case, Dusan's market value is low and his performance score is also lower, so the model may select him to allow space for other more costly and better performing players that would not compromise on the team's overall performance. We appreciated that the model does not function in a silo, it takes into account the entire team's composition when coming up with an optimal solution. We observe that selection of Dusan does improve the objective value to 7,270,951,332.98.

## **7) Which players should we select if we want to use our entire budget and only pick 3 players?**

Selected OptiGuide Response:

*Absolutely!*

*Before, when we didn't limit the number of players chosen, the optimization resulted in an objective value of 7209217116.777735, meaning that this was the best possible team composition, maximizing performance and value while minimizing cost, under the original set of constraints (budget, team needs, one team per player, exclusion and inclusion of certain players).*

*However, after adding the constraint that we could only select 3 players in total, the objective value fell to 6969676070.450805. This decrease in the objective value means that limiting the number of players that can be selected naturally reduces the overall team performance and value, as there are fewer players contributing.*

*This illustrates a key principle of optimization - changing constraints can alter the optimal solution. The challenge is always to find a balance that satisfies all of your objectives and constraints to the best possible degree.*

*So, if your priority is to exactly spend your entire budget and only pick 3 players, the new solution is optimal for you. If your priority is the overall team performance and value, you might want to stick with the first solution.*

Question Intent and Model Performance:

Up until this point, our model optimized by selecting as many players as it could to increase the value as much as possible within the budget constraint. While that could be helpful for teams with a lot of players, most teams probably aren't looking for that many players to

transfer in a single season. More realistically, teams would be looking to consider a more specific number of players, usually between two and five. This question was designed to test how our model performed when asked to use the budget to select a smaller number of players, in this case three.

Under these constraints, the model picked Casemiro, Matthias Ginter, and Mohamed Salah for an optimal value of 6,969,676,070.45., which is less than the optimal value. We have to keep in mind, though, that an assessment of the optimal value depends on the context of what we're trying to achieve. Initially, by giving the model freedom to select the perfect mix of players within a budget allowed the model to balance out performance and economic costs. In this case, the result is optimal for us given that we do want to limit the total number of transfers and spend the entire budget on that limited number. Overall, we were pleased with the model's performance on this question.

**8) Which players should we select if we want to use our entire budget and only pick 3 players and one of them has to be from the Attack position, one of them has to be from the Defender position, and one has to be from the Midfield position?**

Selected OptiGuide Response:

*After optimizing the selection process based on your preference to use your entire budget and pick only 3 players, one from each position (Attack, Defender, Midfield), the algorithm provided an optimal team selection. The selection yielded a total value, which is the function of normalized player performance and their market value, of 6969676070.450805.*

*This result is slightly lower than the original objective value of 7209217116.777735, obtained without the constraints of picking only 3 players and one from each position.*

*This decrease may be due to the fact that the original problem optimization allowed for more flexibility in player selection, which lets the algorithm choose players that offer the best value considering performance and price, regardless of their positions or the number of players to be selected.*

*In contrast, the updated problem, which explicitly enforces the constraint to select players based on positions and limits the selection to only 3 players, narrows down the choice set, which could ultimately yield a lower optimized value.*

Question Intent and Model Performance:

Because our model did so well with the previous question, we asked it to take another constraint into consideration and specified the positions of the players it needed to select. In this case, it selected the same three players - Casemiro, Matthias Ginter, and Mohammed Salah. The optimal value, therefore, remained the same. While this question could be delegitimized as a *true test* of the model, given that the same arrangement of players it previously selected still met the new conditional constraint, it is true that Casemiro is a midfielder, Ginter is a defender, and Salah is an attacker. We consider the model to have done well here.

**9) Which players should we select if we want to use our entire budget and only pick 4 players, 2 from the Attack position and two from the Midfield position?**

Selected OptiGuide Response:

*Sure, let me explain the results.*

*This model was designed to select the best combination of football players to maximize the team's performance given a set amount of budget. It considered each player's market value, normalized performance score, and position (Attack, Midfield, or Defender). The combination of players chosen must also fulfill the team's needs for a specified number of players in each position.*

*The original situation required at least one player from each of the three positions (Attack, Midfield, Defender) to be selected within a budget capped at 150 million euros. The model returned an optimal solution with a total performance score of 7209217116.777735.*

*In contrast, the new request shifted the focus to only two positions - Attack and Midfield. The model was asked to select a total of only 4 players consisting of 2 players from each of these two positions. The budget remained capped at 150 million euros. The model returned an optimal solution with a total performance score of 7178780473.7549.*

*The difference in total performance score between the original solution and the new one is due to the change in the constraints of the model. The new constraints are more restrictive, focusing on fewer positions and limiting the total number of players to be selected. It might have led to the selection of less than optimal players who fit the budget constraints but did not maximize the performance score to the same degree as the original setup.*

*In conclusion, considering these constraints, this combination of players represents the best possible arrangement to achieve maximum performance within the budget.*

Question Intent and Model Performance:

Here, we're pushing our model even further by adding another layer to our previous two questions. When we alter the requirement to select a total of four players (2 attack and 2 midfield), the model performance comes out to be 7,178,780,473.75 which is lower when compared to the original model. Once again, the output provides similar reasoning to the other two related questions. It tells us that our additional restriction makes the model more restrictive and, therefore, decreases its performance potential. The players it selected in this solution were Marten de Roon, César Azpilicueta, Mohamed Salah, and Lionel Messi. The model didn't perform quite as well in this instance. While Marten de Roon is a midfielder and Mohamed Salah and Lionel Messi are attack positions, César Azpilicueta is a defender and, therefore, doesn't meet the constraints.

**10) Which players should we select if we only want players whose total\_goals\_2022 was at least 5?**

### Selected OptiGuide Response:

*By adding a constraint requiring each player to have scored at least 5 goals in total in 2022, the optimized value of the objective function increases from 7209217116.77 to 7237542829.70, indicating that the total performance score multiplied by market value increased by incorporating this constraint.*

*This suggests that players who scored at least 5 total goals in 2022 have higher performance scores and/or market values on average compared to the overall players. Therefore, by narrowing down the player selection to these players, it tends to result in a higher total performance score and market value.*

### Question Intent and Model Performance:

In this question, we asked the model to consider a completely different type of constraint - one that might be valuable if these were to be employed in a real context. There could be performance thresholds that teams want to filter by when considering transfer candidates. In this case, we ask the model to consider only players who scored at least 5 goals in 2022 in the optimal solution. When we asked the model more versions of this question, it could not handle more vague language like “*Which players should we select if we only want players who scored more than 5 goals in 2022?*.” Only by including the column name explicitly were we able to prompt it to generate an output.

When taking the constraint into account, the model optimization value came out to be 7,237,542,829.70, which is higher than the original one. This implies that this mix of players have higher overall performance scores that contribute better to the overall team's success. The players the model selected here were Dusan Tadić, Robert Lewandowski, Jordi Alba, Dani Parejo, Mohamed Salah and Sebastián Coates, all of whom had 5 or more goals in 2022. We consider the model to have performed well on this prompt

## **Appendix 1: Chat Transcript**

We ran our questions through OptiGuide several times to verify consistency and used two different notebooks to do so. The links to our notebooks are below:

[https://github.com/gmoellering/dabp/blob/main/optiguide\\_pt1.ipynb](https://github.com/gmoellering/dabp/blob/main/optiguide_pt1.ipynb)

[https://github.com/gmoellering/dabp/blob/main/optiguide\\_pt2.ipynb](https://github.com/gmoellering/dabp/blob/main/optiguide_pt2.ipynb)