A screenshot of a video game

Description automatically generated

**Project FIFA**

BAMS 508 – Final Group Project

# Executive Summary

Graphical user interface, application

Description automatically generatedFantasy leagues are catching up heat in the recent years. Most fans use their intuition while assembling their fantasy teams because there is no perfect mantra to win. It is also highly dependent on your luck as fine margins decide who wins these leagues. But what if we ask optimization experts with no knowledge of the game to place a team, they see this as a binary integer problem. To help the fans to assemble the perfect FIFA World Cup Fantasy team, our team is building a binary integer optimization model very much like the capital budgeting program we did in class. Our model is a data-driven approach to consider the historical data to create an optimal list of players for your fantasy team.

To summarize, our model tries to maximize the overall points of 11-starting players (plus 4 substitutions) with a 100-million-dollar budget. Each player cost varies price but with different scoring potential, which will be quantified with our predicting model. There are several constraints taken into consideration such as formation and country quota limit which we will discuss further later in this report. The table below show the list of the players that made in our final team.

Table

Description automatically generatedAs you can see in the table below, the final player selected was not all the ones with the highest adjusted score per dollar, which would be an initiative matrix to measure players. In this report, you will find how we create the model, finding the best team for the quarter finals, why some unknown heroes such as Goncalo Ramos (#4) made in the list while big names such as Lionel Messi didn’t make it.

# Introduction

2022 FIFA World Cup is the hot topic right now. With the launch of the official FIFA World Cup Fantasy game, all the fans are finally ready for the big moment after four years of waiting. Our final project will create an optimization model to recommend the best 11-man squad to start, to win the fantasy game. Because of the timing of the FIFA World Cup games, our project will analyse the actual data from all 48 group-stage games + 8 Round of 16 knockout matches, then will have the optimization model ready specifically for the quarterfinals.

For the Fantasy game itself, the purpose is to score as many points as possible with a 11-player starters squad and you are allowed to keep 4 substitutes as reserves in case things didn’t go your way. To assemble the squad, you have $100 million budget, and each player will have a price tag. There is a total pool of 832 player for the group stage and only 416 players will play at the round of 16. Please see below screenshot of the actual game for an example of the team creation.

Graphical user interface, application

Description automatically generated

To score points, the game will take each player’s actual game stats into account and calculate the score. For example, if a Forward played 75 minutes (+1), 3 shots on target (+2), score two goals (+10), one assist (+3), and one yellow card (-1), he will score a 15 total points.

Shown below is the **scoring system** to allocate points to each player depending on the action. All these points are then aggregated after every match to get the total score for that match for the player.

Table

Description automatically generated with medium confidence

For a complete set of rules for scoring and the Fantasy game in general, please refer to the official game guidelines below:

FIFA Fantasy Classic - Game Guidelines: <https://play.fifa.com/fantasy-classic/help/guidelines>

# Optimization Formulation

## Decision Variables

Each player from the top 8 teams, which have qualified for quarter finals, will be our decision variable for our model.

Player 1 = X1, Player 2 = X2 and so on.

Each player will be a binary decision variable.

**Xi =** 0 if the player is not selected in the optimal team for quarter finals

1 if the player is selected in the optimal team for quarter finals

Here, i ranges from 1 to 139.

We took only those players from the top 8 teams who had a Total FIFA Score > 0

## Objective Coefficients

We calculated an adjusted weighted score for each player. This score was used as the coefficient for their respective decision variables.

p = weight given to the score of Match 4

M1i = Score of the player i for Match 1 (Group Stage)

M2i = Score of the player i for Match 2 (Group Stage)

M3i = Score of the player i for Match 3 (Group Stage)

M4i = Score of the player i for Match 4 (Round of 16)

We take the average score of all group stage matches (1,2and 3) and multiply it by the weight of group stage. We multiply the weight of the knock-out stage match (4) and add both the weighted scores to get the final weighted score.

We gave the Match 4 score a higher weightage as it was a knock-out round and was more recent.

## Objective Functions

We aim to maximize the total adjusted weighted score for our team by selecting the most optimal players. To achieve this, we maximize -:

## Constraints

1. **Total players Constraint:** The primary hard constraint that we have is that our fantasy team can only comprise 15 players.
2. **Budget Constraint**: There is a budget cap of $100 million and each player has a price tag
3. **Position Constraints**: You are allowed to have at max 3 forwards, 5 midfielders, and defensive players. There will be at least one goalkeeper as well in the team. These are all the possible formations that a playing 11 can have:

4-4-2

4-3-3

4-5-1

3-4-3

3-5-2

5-2-3

5-3-2

5-4-1

Apart from this, we also need to choose our 4 substitutes, one of which must be a goalkeeper. Although points from these players do not contribute to our fantasy points, these players have a price tag and must follow the total budget constraint.

1. **National Team Constraint**: You are allowed to choose 5 players max per nation for the quarterfinals.
2. **Yellow Card Constraint**: A player who got 2 yellow cards in consecutive games can’t play the next game
3. **Red Card Constraint**: A player who gets a red card cannot play the next game.

# Optimization Analysis

## Data Collection

Shown below is our data schema, which consists of three data sources (tables): Players, Player Stats, and Match. The symbol in the shape of an oval represents the attributes for each data source. The relationships between each of the data sources are depicted by the diamond symbols. This is the data schema we developed to describe the relationships between the data sources we gathered from the website.

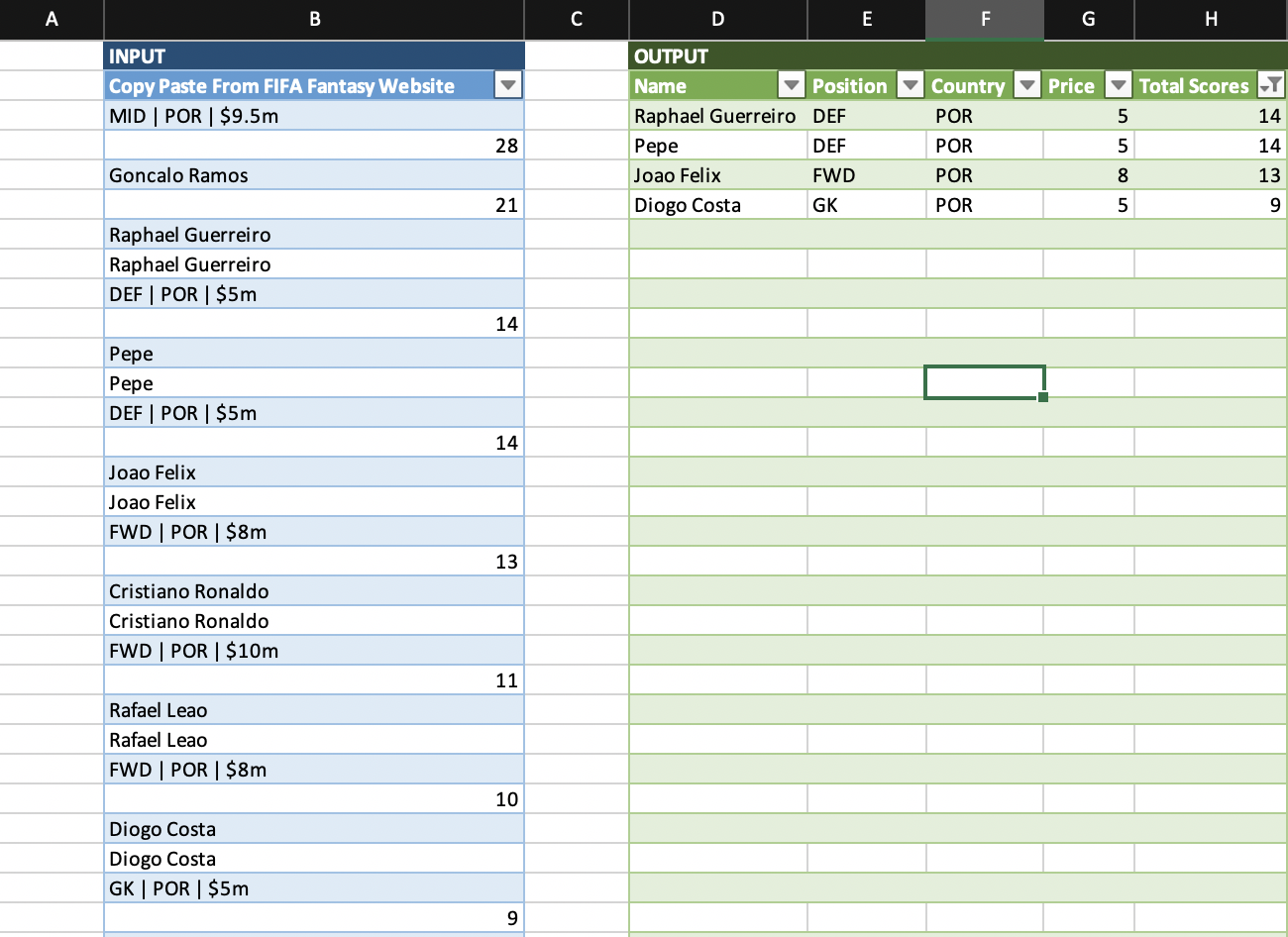
Diagram

Description automatically generated

We created an excel template using Power Query, which takes a certain text input from our player pool database and gives an output in the desired format (player name, country, position, and budget). Shown below is what our input and output look like, we select and copy the players data.

Graphical user interface, application, website

Description automatically generated



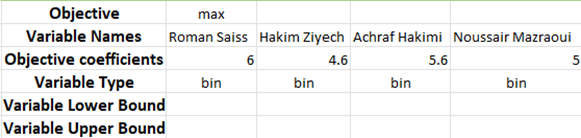
After selecting and copying our data from the Fantasy League (as shown in the previous image), we paste the text component of it in the INPUT section (highlighted in blue in the above image). After filling in our input this way, we refresh the OUTPUT table (highlighted in green in the above image), which gives us some of the important data points – Name, Position, Country, Price, and Total Scores. Using this approach would allow us to quickly fetch the player information for all the 139 players (8 teams).

So far, we have collected player information for all 139 players. Other features that also must be fetched are – yellow cards, red cards, assists, and goals scored. Such data points were manually inserted by us for every player as the API was not publicly available (limitation) which would have allowed us to fetch everything through a JSON query.



## Data Modeling

After collecting the data for all 139 players, we have most of the attributes to populate the Gurobi input sheet. We are using players’ names as variable names because each of them is unique. Below is a screenshot of the decision variable and objective section of the Excel to Gurobi input sheet.



To calculate the objective coefficients which are the adjusted scores for each player. We used the equation mentioned in the previous section and set p% at 40% for the one round of 16 match, which would also give 60% weight for the total scores for the three group stage matches (20% per match). We gave the Match 4 score a higher weight percentage because it was in the knock-out round, which is more recent, reflects the players’ condition better, and more relevant, reflects the quarter finals match intensity better.

Below are a list of players’ Adjusted Average Scores to be used as variable coefficients in the model.

Graphical user interface, table

Description automatically generated

Then we set up the constraints following the logic in the previous section to find the 15-man squad. The total player constraint, budget constraint, position constraint, national team constraint, yellow card constraint, and red card constraint can be formulated in Excel to Gurobi model like below.

Graphical user interface, application, table

Description automatically generated

Once we have found the best 15-man squad, we need to find out which 11 players we need to start because only the starters’ scores are counted in the game. To find the best 11-player starters, we loosen some of the constraints from the above 15-man squad case according to the methodology listed out in the constraint section before. One thing to mention is because we are now finding 11 players instead of 15, we must decrease the budgetary cap to reflect the change. We have assumed that we are allocating 85% of the total $100 budget towards starts. The new constraints are represented in the excel below.

A picture containing chart

Description automatically generated

## Results & Scenarios

With the Excel to Gurobi file set up, we pulled the optimal solutions for the best 15-man squad and the 11-man starters shown as below. We can interpret the results as by picking the 11-man starters (**Bruno Fernandes, Bukayo Saka, Daley Blind, Denzel Dumfries, Goncalo Ramos, Jordan Pickford, Kylian Mbappe, Pepe, Philip Foden, Raphael Guerreiro, and Richarlison**), you fantasy team will have a optimal maximum expected score of 83.8 total points. Please note that the optimal maximum objective value for the 15-man squad does not provide as much value because the game only recognizes the scores for the 11-man starters.

Best 15-man Squad Gurobi Result

Text

Description automatically generated

Best 11-man Starter Gurobi Result

Text

Description automatically generated

Because we used two assumptions (40% round-of-16 score, 85% allocation for starters) when setting up the model, we tweaked our optimization model to see variations of different teams depending on how we might want them.

Table

Description automatically generated

First, we gave different budget weightage for our playing 11 and 4 substitutes. We allocated 85 million out of 100 for our top 11 and got the list. An interesting insight we got here was, even after changing this budget weightage for our playing 11 to 80 and 90, we got the same list of players. This implied that our budget allocation within this range had no effect on our list of top 11 players.

Next, we tried to give weightage of points from the first 3 group stage matches and the last knockout match. Tweaking our weights makes sense here as some people might want to select a team based on the overall form of a player throughout the tournament, while some might want to give preference to the form of players in the last knockout match itself. Therefore, we assigned 30, 40, and 50 percent weightage to the points scored during the knockout Round of 16 game and got the list. We observed that our team lists remain the same for 40 and 50 percent while differed for the 30 percent weights. This is a great insight as these weights accommodate the subjective preference based on which you would want the list of your team.

Table

Description automatically generated

# Conclusion & Insight

Selecting an optimal fantasy team is a great binary integer problem. The problem at hand is to either choose or reject a player for our team given several constraints that we must consider. Below are some of the interesting insights we have noticed.

## Defender – Big Risk, Big Rewards

When teams don’t concede any goals or keep a clean sheet, all the defenders get 5 points. On top of that, if a defender scores, especially Centre Backs, it makes their overall points shoot up. Therefore, some defenders might be picked up by our model even though there is a rare chance they will be able to replicate the same performance. An instance of this is Pepe being selected by our model, just because he scored a goal in the Round of 16 and has a smaller price tag.

## Unknown Hero

Football can get very unpredictable at times, which a binary integer model can’t account for. The best example to illustrate this is Goncalo Ramos of Portugal who started his first-ever world cup match for Portugal against Switzerland. He went on to score a hat-trick, which was also the first hat-trick of the tournament. This got him a lot of points and given his low-price tag, he is an easy pick for our model. However, him starting the next game is uncertain as it’s not every day you get to replace the GOAT Cristiano Ronaldo.

## The Notable Left outs

If you leave Ronaldo, Messi, and Neymar from your team, there are a lot of people who would question that. There is no doubt these players are great. However, there’s no model or team that can have these players together, given how expensive they are. So, in order to justify their price tag, their stats must be insanely good to be picked up by our model. Kylian Mbappe, who is another expensive and high-profile player is an example whose score/price is high even after its high valuation and gets picked up by our model.

# Others

## Limitations & Improvements

* Data collection using an API. Finding the player's information (budget, country, position, etc.) through the FIFA Fantasy League API proved difficult because it is not made available to the public. This would have made our data collection process easier, as we could have collected the players' data using Python JSON tools. Shown below is the error we faced:

['Bad', 'request', 'Bad', 'request', 'A', 'critical', 'request', 'has', 'been', 'detected', 'and', 'therefore', 'blocked.', 'IP:', '128.189.82.168', 'Reference:, '917283553972466751310588293993578027401', 'Please', 'report', 'the', 'incident', 'including', 'request', 'ID', 'to', 'FIFA', 'Service', 'Desk.', 'Thank', 'you!']

* FIFA website took down all the players at the last minute who are not in the quarter finals. We were initially modelling the problem for 16 teams (~ 800 players), but we had to change our scope at the last minute, and we had to resort to 8 teams (139 players).
* In our analysis, the sample size per player is limited to 4 matches, but ideally, we would like to have more matches and more data points to predict a more accurate score that reflect each player’s real scoring potential.

## Future Applications

* During the project, we noticed that the FIFA fantasy game itself have no auto-select function for the game after group stage. Our optimization model could potentially be used to create such feature for the platform in the future.
* The approach taken for this project will work better for Fantasy Leagues restricted to country club-level leagues (Premier League, La Liga, etc.), as a lot of the information on Players and Matches are readily available using API’s and Kaggle site.
* Additional analysis, analyze team strength, schedule strength, team chemistry – these are some of the data points that would’ve helped us to gain more insights and improve the prediction accuracy regarding the player pool data we possess. Some of the important ones being – offense and defense rating, pace, strength, and skills rating.

# Citation

FIFA World Cup Qatar 2022. (2022, November 14). *FIFA World Cup™ Fantasy*. FIFA+ Crypto.com Play Zone. Retrieved December 9, 2022, from https://play.fifa.com/fantasy-classic/help/guidelines

# Appendices

## Chart, histogram Description automatically generatedChart, bar chart, histogram Description automatically generatedAdjusted Score per Dollar Chart for Eight Quarter Final Teams

## Top 50 Players Order by Adjusted Score per Dollar

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rank** | **Name** | **Position** | **Country** | **Price** | **Adjusted Average Score** | **Adjusted Score/ Dollar** | **Top 15** | **Top 11** |
| 1 | Denzel Dumfries | DEF | NED | 6 | 8.2 | 1.37 | 1 | 0 |
| 2 | Pepe | DEF | POR | 5 | 6.8 | 1.36 | 1 | 1 |
| 3 | Roman Saiss | DEF | MAR | 4.5 | 6 | 1.33 | 1 | 0 |
| 4 | Goncalo Ramos | FWD | POR | 6.5 | 8.4 | 1.29 | 1 | 1 |
| 5 | Daley Blind | DEF | NED | 5.5 | 7 | 1.27 | 1 | 1 |
| 6 | Raphael Guerreiro | DEF | POR | 5 | 6.2 | 1.24 | 1 | 1 |
| 7 | Dejan Lovren | DEF | CRO | 4.5 | 5.2 | 1.16 | 1 | 0 |
| 8 | Nayef Aguerd | DEF | MAR | 4 | 4.6 | 1.15 | 0 | 0 |
| 9 | Bukayo Saka | MID | ENG | 8 | 9.1 | 1.14 | 1 | 1 |
| 10 | Achraf Hakimi | DEF | MAR | 5 | 5.6 | 1.12 | 0 | 0 |
| 11 | Danilo | DEF | BRA | 5.5 | 6 | 1.09 | 0 | 0 |
| 12 | Luke Shaw | DEF | ENG | 5 | 5.4 | 1.08 | 0 | 0 |
| 13 | Monir El Kajoui | GK | MAR | 4 | 4.2 | 1.05 | 0 | 0 |
| 14 | Harry Maguire | DEF | ENG | 5.5 | 5.6 | 1.02 | 0 | 0 |
| 15 | Richarlison | FWD | BRA | 7.5 | 7.5 | 1.00 | 1 | 1 |
| 16 | Noussair Mazraoui | DEF | MAR | 5 | 5 | 1.00 | 0 | 0 |
| 17 | Yassine Bounou | GK | MAR | 4.5 | 4.5 | 1.00 | 1 | 0 |
| 18 | Bruno Fernandes | MID | POR | 9.5 | 8.9 | 0.94 | 1 | 1 |
| 19 | Kylian Mbappe | FWD | FRA | 11.5 | 10.6 | 0.92 | 1 | 1 |
| 20 | Andries Noppert | GK | NED | 4.5 | 4 | 0.89 | 0 | 0 |
| 21 | Jordan Pickford | GK | ENG | 5.5 | 4.8 | 0.87 | 1 | 1 |
| 22 | Thiago Silva | DEF | BRA | 6 | 5.2 | 0.87 | 0 | 0 |
| 23 | Theo Hernandez | DEF | FRA | 5 | 4.3 | 0.86 | 0 | 0 |
| 24 | Olivier Giroud | FWD | FRA | 7.5 | 6.3 | 0.84 | 0 | 0 |
| 25 | Eder Militao | DEF | BRA | 5 | 4.2 | 0.84 | 0 | 0 |
| 26 | Lovro Majer | MID | CRO | 5 | 4.2 | 0.84 | 0 | 0 |
| 27 | John Stones | DEF | ENG | 5.5 | 4.6 | 0.84 | 0 | 0 |
| 28 | Joao Felix | FWD | POR | 8 | 6.4 | 0.80 | 0 | 0 |
| 29 | Nicolas Otamendi | DEF | ARG | 5 | 4 | 0.80 | 0 | 0 |
| 30 | Marcos Acuna | DEF | ARG | 4.5 | 3.4 | 0.76 | 0 | 0 |
| 31 | Josko Gvardiol | DEF | CRO | 4 | 3 | 0.75 | 0 | 0 |
| 32 | Ivan Perisic | MID | CRO | 7.5 | 5.6 | 0.75 | 1 | 1 |
| 33 | Borna Sosa | DEF | CRO | 3.5 | 2.6 | 0.74 | 0 | 0 |
| 34 | Philip Foden | MID | ENG | 8.5 | 6.3 | 0.74 | 1 | 1 |
| 35 | Casemiro | MID | BRA | 6 | 4.4 | 0.73 | 0 | 0 |
| 36 | Alisson | GK | BRA | 6 | 4.4 | 0.73 | 0 | 0 |
| 37 | Dominik Livakovic | GK | CRO | 5 | 3.6 | 0.72 | 0 | 0 |
| 38 | Alex Sandro | DEF | BRA | 5 | 3.6 | 0.72 | 0 | 0 |
| 39 | Josip Juranovic | DEF | CRO | 5 | 3.6 | 0.72 | 0 | 0 |
| 40 | Jude Bellingham | MID | ENG | 7.5 | 5.2 | 0.69 | 0 | 0 |
| 41 | Lionel Messi | FWD | ARG | 10.5 | 7 | 0.67 | 0 | 0 |
| 42 | Hakim Ziyech | MID | MAR | 7 | 4.6 | 0.66 | 0 | 0 |
| 43 | Jordan Henderson | MID | ENG | 6.5 | 4.2 | 0.65 | 0 | 0 |
| 44 | Ricardo Horta | FWD | POR | 6.5 | 4.2 | 0.65 | 0 | 0 |
| 45 | Lucas Paqueta | MID | BRA | 7 | 4.5 | 0.64 | 0 | 0 |
| 46 | Rafael Leao | FWD | POR | 8 | 5 | 0.63 | 0 | 0 |
| 47 | Diogo Dalot | DEF | POR | 5.5 | 3.4 | 0.62 | 0 | 0 |
| 48 | Ruben Dias | DEF | POR | 5.5 | 3.4 | 0.62 | 0 | 0 |
| 49 | Alexis Mac Allister | MID | ARG | 7 | 4.3 | 0.61 | 0 | 0 |
| 50 | Gonzalo Montiel | DEF | ARG | 5 | 3 | 0.60 | 0 | 0 |