Applying Optimization to NBA Lineups

Cameron Tomko

OIE 559

Professor Andrew Trapp

Submitted: October 11, 2023

Table of Contents

Abstract

In this project, I aimed to create a roster for the 2023 – 2024 National Basketball Association (NBA) season that optimized win share (WS), which calculates a player's total impact of winning. Using data collected from Basketball Reference for all available NBA players from the 2022 – 2023 season, the model implemented binary integer optimization and satisfies constraints to ensure the realistic feasibility of the team while staying below the 2023 – 2024 NBA salary cap. The decision variable results of the model indicate players who were selected for the roster. Simulations of the rosters were conducted using WhatIfSports. Simulating sets of games and recording game statistics show the impact of optimization and win share as a viable metric to assess influence on winning in the NBA. Through the creation of the binary integer model and the results of the simulations, the report provides a captivating argument to showcase the effectiveness of optimization in sports analytics.

Executive Summary

In recent memory, there has been constant growth around the importance of collecting data and utilizing it to help drive decision-making in every industry. The effects of data and its impact have been transferred into the professional sports industry as teams have implemented new tools to understand data. The National Basketball Association (NBA) currently uses analytics and data to form rosters in hopes of winning the NBA championship. The project's main objective is to illustrate the solving power of binary integer programming and its implementation into the NBA. The model was formulated to optimize the total win share value to create a roster that can be successful against all-time NBA teams. Win share is a numerical advanced statistic that represents a player's ability to influence winning.

Data collection was the first step in further understanding the task at hand. Data from Basketball Reference was gathered, focusing on player name, age, salary, win share, and the number of games started. These data fields built the model with various constraints that improved the feasibility of the results. The model selected was a binary integer programming model. Creating a model has various components including identifying the sets, setting an objective function, which is either a maximized or minimized function, and selecting decision variables and constraints, which the model satisfies to produce an optimal solution.

The goal of this binary integer program is to maximize the sum-product of the win share for a player and the associated variable. The project deployed an iterative approach to creating a finalized model with each instance creating more realistic results. The finalized model has constraints, which bind the average age, cap the max salary, and limit one starter caliber player at each position. The model was created and optimized using the Gurobipy library in Python. The optimized results for the first model were 84.8 and 76.2 for the final model. The results of the selected decision variables for the first and final model were imported into WhatIfSports, which performs game simulations for the NBA. The created teams played against the 2015 – 2016 Golden State Warriors and the 2022 – 2023 Denver Nuggets. The simulation results provide evidence that indicates the success of the binary integer models as the overall winning percentages for a 60-game span were 0.867 and 0.682 for the two models, respectively.

1. Introduction

The world around us is surrounded by hundreds of millions of data points. Using data has become a focal point in many industries. From trying to understand what data is being gathered to gaining insights on the data and applying them to a business problem, data drives decision-making. This will only increase as more companies begin to deploy techniques to wrangle and analyze data. The same can be said about the sports industry as data and analytics have had a tremendous impact on team management and in-game decisions. Sports analytics first became important after Major League Baseball's Oakland Athletics' general manager utilized data and analytics to create a team that performed at a high level while maintaining a salary below the league average. Since then, the use of analytics in sports has taken off as almost every movement a player makes is being recorded and used to calculate a metric.

In this project, I looked to explore binary integer programming in the National Basketball Association. The idea of the project originates from a research paper from Stanford University, where the researcher used integer programming to draft a Major League Baseball team based on different constraints that make up a team (Holt, n.d.). In this project, I looked to maximize the win share of a team with current-day NBA players for the 2023 – 2024 NBA season. Win share (WS) is a metric that calculates a player's impact on helping a team win (*NBA Win Shares*, n.d.). By adding various constraints to create the team and remain within the salary cap for the season, I attempted to create a realistic team that could be used to simulate games. Through using WhatIfSports, I simulated games with the different teams and recorded the results of the games. Between iterative modeling and simulation with different teams, the paper helps demonstrate that binary integer programming can help business stakeholders simulate different scenarios and gain valuable insights.

2. Background

The background section of the paper provides information on the main focuses of the business problem of the project. A strong understanding of the interworking of the National Basketball Association and the use of analytics in the NBA is crucial to fully grasp the content discussed throughout the paper. The background section aims to provide this understanding.

2.1 National Basketball Association (NBA)

The National Basketball Association, more commonly known as the NBA, was founded in 1946 as the Basketball Association of America. Soon after its founding, the league merged with the National Basketball League in 1949 and the NBA was born (*History*, n.d.). The NBA currently hosts 30 teams across two different conferences, the Eastern and Western conferences. Within each conference, there are three different divisions, each having five teams. All of the NBA teams are located in North America, with 29 being scattered across the continental United States and the remaining one located in Toronto, Canada. The NBA showcases a global platform as in the 2021 – 2022 season had over 140 players from 40 international countries (*Our Leagues*, n.d.). The goal that every team sets out to accomplish yearly is to compete for and ultimately win the NBA Championship. Before getting to the final goal, teams must play an 82-game schedule to gain a chance to be crowned the ultimate champions.

2.2 Sports Analytics and Statistics in the NBA

As mentioned, analytics and data have had a growing influence on sports. There is an annually hosted analytics conference for Wharton's People Analytics Conference, where NBA Commissioner Adam Silver discussed the league's use of analytics in managing a player throughout the season to promote a healthy and extensive career. The NBA tracks a player's every move every time they step on a court for games and even practice. Even more the NBA is tracking sleep data, tracking diet, and taking saliva samples to gain a further explanation of why players get tired and how to prevent this. The league uses analysis to help determine how many minutes a player should play, who they will guard, and when in the game they should be substituted out ("The NBA's Adam Silver," n.d.). Furthermore, more relevant to this project, the NBA uses analysis for advanced scouting to help determine fits for their current roster by evaluating possible free-agent acquisitions or draft-eligible players (Editor, 2018).

3. Methodology

3.1 Data Collection

Before formulating the optimization model, I collected data so the model could be optimized using real NBA player data. Collecting data allowed for the optimized models and the results to have the highest accuracy. Before gathering the data, I needed to decide which fields would be needed for the model. In the first iteration of the model, I determined that the following attributes were necessary for the model: player name, position, win share, salary, and unique player key. The data was gathered using basketball-reference.com for all the data. The win share data is from the 2022 – 2023 NBA season. (*2022-23 NBA Player Stats: Per Game | Basketball-Reference.Com*, n.d.). The salary data is for the 2023 – 2024 season as the optimization would be happening for next year's NBA season. The player statistics and salary data were originally presented as two separate datasets. The two datasets were joined on the unique player keys as the datasets had the same primary key. The combined dataset created the data needed for the first and second iterations of the mathematical model.

For the final iteration of the mathematical model, I realized more data would be needed to obtain a more realistic construction of an NBA roster. The data collection for this iteration contained the same data from the first iteration with the addition of player age, number of games started, and number of games played. Each player's data for the number of games started was divided by the number of games played to create an aggregate attribute called 'Percentage of Games Started'. The data which created this new field was based on the 2022 – 2023 NBA season. In a similar process as previously mentioned, the new fields were joined to the first model's dataset on the unique player key field. The dataset used for the final mathematical model had the following attributes: player name, position, age, win share, salary, percentage of games started, and a unique player key. The 'OIE_Project_Final_Data.csv file attached to the submission provides the dataset that the model uses. An example of the data is shown below in Figure 1.

	A	B	C	D	E	F	G	H	I
1	Unique Identifier	Player	Position	Age	WS	Salary	GS_Percent	G	All-Star
2	jokicni01	Nikola Jokić	C	27	14.9	47607350	1	69	1
3	sabondo01	Domantas Sabonis	C	26	12.6	30600000	1	79	1
4	butleji01	Jimmy Butler	PF	33	12.3	45183960	1	64	1
5	embiijo01	Joel Embiid	C	28	12.3	47607350	1	66	1
6	gilgesh01	Shai Gilgeous-Alexander	PG	24	11.4	33386850	1	68	1
7	tatumja01	Jayson Tatum	SF	24	10.5	32600060	1	74	1
8	doncilu01	Luka Dončić	PG	23	10.2	40064220	1	66	1
9	allenja01	Jarrett Allen	C	24	9.5	20000000	1	68	1
10	claxtni01	Nic Claxton	C	23	9.2	9625000	1	76	0
11	davisan02	Anthony Davis	C	29	9	40600080	0.964285714	56	1
12	lillada01	Damian Lillard	PG	32	9	45640084	1	58	1
13	mitchdo01	Donovan Mitchell	SG	26	8.9	33162030	1	68	1
14	brunsja01	Jalen Brunson	PG	26	8.7	26346666	1	68	0
15	looneke01	Kevon Looney	C	26	8.7	7500000	0.853658537	82	0
16	antetgi01	Giannis Antetokounmpo	PF	28	8.6	45640084	1	63	1
17	derozde01	DeMar DeRozan	SF	33	8.5	28600000	1	74	1
18	mobleev01	Evan Mobley	PF	21	8.5	8882640	1	79	1
19	hardeja01	James Harden	PG	33	8.4	35640000	1	58	1
20	vucevni01	Nikola Vučević	C	32	8.3	18518519	1	82	1
21	markkla01	Lauri Markkanen	PF	25	8.2	17259999	1	66	1
22	randlju01	Julius Randle	PF	28	8.1	28226880	1	77	1
23	lopezbr01	Brook Lopez	C	34	8	25000000	1	78	1
24	plumlma01	Mason Plumlee	C	32	7.9	5000000	0.759493671	79	0
25	siakapa01	Pascal Siakam	PF	28	7.8	37893408	1	71	1
26	goberru01	Rudy Gobert	C	30	7.8	41000000	1	70	1
27	curryst01	Stephen Curry	PG	34	7.8	51915615	1	56	1

Figure 1: Sample Dataset for Model

3.2 Formulating the Model

3.2.1 Set Definition

P: Set of players, indexed by i, with n total

C: Set of positions, indexed by j, with m total

G: Set of players who can start, indexed by k, with l total

3.2.2 Decision Variable Definitions

$$x_i = \begin{cases} 1 \; if \; the \; player \; is \; selected \\ 0 \; otherwise \end{cases}$$

3.2.3 Parameter Definitions

w_i is the number of win shares associated with $i \in P$

s_i is the salary associated with $i \in P$

a_i is the age associated with $i \in P$

b_i is if an all-star selection was awarded to $i \in P$

3.2.4 Assumptions

There are assumptions for the collected data and the binary integer programming model. Having the following assumptions allows for planning for the next NBA season and will account for all players who will play.

1. Assume that all the data I have collected from Basketball Reference is accurate and correct.
2. Assume all NBA players in the dataset will play during the 2023 – 2024 season.
3. Assume all eligible players for the 2023 – 2024 season are represented in the dataset.
4. Assume a player having an all-star selection means that a player has been selected as an all-star once in their career.
5. Assume there is no lower bound for the salary cap.

3.3 Iterative Mathematical Model Creation

For deploying a binary integer programming model, I undertook an iterative approach in hopes of returning an NBA lineup that resembles a realistic roster. By conducting an iterative approach, I attempted three different models and updated the objective functions and constraints as needed. The next subsections discuss the modeling of the program.

3.3.1 First Mathematical Model

In the first approach of handling this problem, the objective function is to maximize the sum-product of the matching win share and the binary variable for each player. By maximizing this, the optimized model will return a total win share score for the team and the corresponding variables that are selected. The objective function is stated below.

$$Maximize \sum_{i \in P} w_i x_i$$

The constraints of the model are listed below:

1. The team has selected exactly ten players:

   $\sum_{i \in P} x_i = 10$
2. The team's salary can be at most \$136.021 million:

   $\sum_{i \in P} s_i x_i \geq 136.021 \; million$
3. The team can only have two players at each position:

   $\sum_{i \in P} x_{ij} = 2 \;\; \forall \, j = 1, \dots, m$
4. Decision variable must be binary:

   $x_i \; \in \{0,1\} \, \forall \, i \in P$

The first constraint indicates that ten players must be selected for the model. The number of selected variables is ten as that is the normal size of an NBA lineup, having five spots for starting players and five spots for bench players. The second constraint indicates that the salary of the team cannot exceed \$136,021,000 as this is the salary cap for the 2023 – 2024 NBA season

(*NBA Sets Salary Cap at $136 Million for 2023-24 Season | NBA.Com*, n.d.). The third constraint tells the model that for the five positions on a basketball team, point guard, shooting guard, small forward, power forward, and center, two players must be selected who play that position. The final constraint ensures that the player selection variable is binary since a player will or will not get selected for the team.

3.3.2 Second Mathematical Model

Having optimized the first model, I attempted a new iteration to minimize the salary of the roster by setting a constraint to keep the maximized win share value. The new objective function is shown in the following:

$$Minimize \sum_{i \in P} s_i x_i$$

The constraints of the second model are shown below. Constraint #3 holds the maximized win share value that was found in the first model:

1. The team has selected exactly ten players:

   $\sum_{i \in P} x_i = 10$

2. The team can only have two players at each position:

   $\sum_{i \in P} x_{ij} = 2 \;\; \forall \, j = 1, \ldots, m$

3. The team must have a total win share value of 84.8:

   $\sum_{i \in P} w_i x_i = 84.8$

4. Decision variable must be binary:

   $x_i \in \{0,1\} \; \forall \, i \in P$

Constraints 1, 2, and 4 were discussed in the previous iteration, which is shown above. The third constraint indicates that the model must select players whose win share sums to the optimized maximum value of 84.8, which was found out through the previous iteration.

3.3.3 Finalized Mathematical

After analyzing the results of the first and second models, with both having the same output, I noticed that the roster consisted of players who were impactful players on their NBA team, but I saw this roster as very unrealistic to have as a current NBA team. The team consisted of 5 all-star players, where the previous highest number of all-stars for an NBA team was four players (*Thompson, Green among Reserves for 2018 NBA All-Star Game*, n.d.). Next, I realized that the roster consisted of nine out of the ten players who started more than half of the NBA games they participated in during the 2022 – 2023 season. For these reasons, I identified the shortcomings of the optimized roster for the first model generation. The unrealistic features helped identify new constraints that would be necessary for a more realistic model. The new model is shown below. This model is the final iteration of the model development.

$$Maximize \sum_{i=1}^{n} w_i x_i$$

The constraints of the model are listed below:

1. The team has selected exactly ten players:

$\sum_{i \in P} x_i = 10$

2. The team's salary can be at most \$136.021 million:

   $\sum_{i \in P} s_i x_i \geq 136.021\ million$

3. The team can only have two players at each position, i.e. point guard, shooting guard, small forward, power forward, and center:

   $\sum_{i \in P} x_{ij} = 2\ \ \forall\, j = PG, SG, SF, PF, C$

4. The team must have an average age above 23.14 years:

   $\frac{1}{10} \sum_{i \in P} a_i x_i \geq 23.14$

5. The team must have an average age below 29.47 years:

   $\frac{1}{10} \sum_{i \in P} a_i x_i \leq 29.47$

6. The team can have at most one player at each position who started over 50% of games:

   $\sum_{k \in G} x_{kj} \leq 1\ \forall j = 1, \ldots, m$

7. Decision variable must be binary:

   $x_i\ \in \{0,1\}\ \forall\ i \in P$

To further understand the constraints, I will address each one. Constraints 1, 2, 3, and 7 are from the first model, which have been discussed. Constraints 4 and 5 work as an upper and lower bound for the average age of the team. These bounds were derived by researching the age of the youngest and oldest NBA teams during the 2022 – 2023 season (*NBA Roster Survey: Facts to Know for the 2022-23 Season | NBA.Com*, n.d.). The last constraint to talk about is Constraint 6. It tells the model that, based on data of players who started 50 % or more of the games they participated in, there is a limit of at most one starter at each position. This constraint was constructed to create a more realistic scenario and ensure there are no starting caliber players on a team's bench rotation.

3.4 Solving the Mathematical Models

To solve the mathematical models, I utilized the Gurobipy package in Python. I created the models by importing the dataset, setting the objective functions, adding the constraints to the model, and optimizing the model. Three Python files are attached to the submission. Each file contains a specific model and is commented thoroughly to ensure the user will be able to understand the steps taken to obtain the solutions to each of the optimized models. Figure 2 is a snippet of the Python code used for building the final mathematical model.

```python
from gurobipy import *
import gurobipy as gp
import pandas as pd
```

```python
# importing dataset
nba_data = pd.read_csv('OIE_Project_Final_Data.csv')
```

```python
nba_data['Salary'] = nba_data['Salary'].astype('int64')
data_types = nba_data.dtypes
```

```python
# objective function coefficients
max_ws_fun = nba_data['WS']
```

```python
# creating model
model  = Model()
```

```
Restricted license - for non-production use only - expires 2024-10-28
```

```python
# length of decision variables
num_vars = len(max_ws_fun)
```

```python
# adding decision variables to model
vars = model.addVars(num_vars, vtype=gurobipy.GRB.BINARY, name='x')
```

```python
# setting objective function for maximizing the number of win shares
model.setObjective(gurobipy.quicksum(max_ws_fun[i] * vars[i]
                                     for i in range(num_vars)), sense=gurobipy.GRB.MAXIMIZE)
```

```python
# setting the max number of players on the team
max_players = 10
model.addConstr(gurobipy.quicksum(vars[i] for i in range(num_vars)) == max_players)
```

```
<gurobi.Constr *Awaiting Model Update*>
```

```python
player_salary = nba_data['Salary']
salary_cap = 136021000
team_salary = gurobipy.quicksum(player_salary[i] * vars[i] for i in range(num_vars))
model.addConstr(team_salary <= salary_cap)
```

Figure 2: Sample Gurobipy Code for Final Mathematical Model

The code shows the process of calling in the necessary libraries and data file, checking the data types, and creating a model. Once the model was created, I began focusing on setting the objective function to maximize the sum-product of the win share and the binary variable for each player. After this step, I go on to add two constraints: setting the number of selections to ten and the sum-product of the salary and variables for each player must not exceed the salary cap of 136.021 million. The rest of the code of the final mathematical model, along with the first two iterations, can be seen in the Python files attached to the submission.

3.5 Simulation of Models

The solution of the optimized models will not only output the maximized objective function of win share but the corresponding variables, which represent players, that create the maximized objective function. By looking at the selected variables, I created a roster to simulate games with. Each roster for the iterations of the mathematical models was imported into WhatIfSports to run simulated games. WhatIfSports.com is a sports simulation website that allows users to create their teams or use past and present teams to simulate the result of a game

(*SimMatchup Basketball - Free NBA Matchup and Basketball Sim Games*, n.d.). Having imported the optimized rosters, I simulated two 30-game sets against two different teams, where the optimized roster will be the home team for 15 games and the away team for the other 15 games. The teams, which the optimized roster played against, were the 2015 – 2016 Golden State Warriors and the 2022 – 2023 Denver Nuggets. The 2015 – 2016 Golden State Warriors are widely known as one of the best teams of all time as they currently hold the best single-season record of all time at 73 wins and 9 losses, a 0.890 winning percentage (*2015-16 Golden State Warriors Schedule | Basketball-Reference.Com*, n.d.). The 2022-2023 Denver Nuggets were chosen as this team won the NBA Championship during the 2022 – 2023 season (*2022-23 Denver Nuggets Roster and Stats | Basketball-Reference.Com*, n.d.).The simulation tests the model's roster against current day and all-time NBA teams with championship level success. When simulating the games, I manually set the minutes per player to ensure the starters were playing the majority of the minutes. For all teams including the modeled rosters and the historical teams, I normalized the minutes to give 32 minutes per game to the five starters and 16 minutes per game to the five bench players. The results for each of the games were recorded, along with the corresponding scores and home team and away team for each game.

4. Results

The next section summarizes the results of the optimization models and simulation of the output rosters. The section includes a discussion on the effectiveness of the additional constraints in hopes of creating more realistic NBA lineups.

4.1 Assessing the Optimization Models

To assess the optimization models, I first began by looking into the roster construction that the model created. As the models were created using binary variables to represent the players selected by the different models, I compiled the variable results and recorded the variable's corresponding player. For the first mathematical model, the goal was to maximize the amount of win share for the roster. This model had fewer constraints than the final model. The roster construction results are referred to as 'Team #1' and are shown below in Table 1. Team #1 had a total of 84.8-win share for the entire roster. The team consisted of five all-stars from the 2022 – 2023 season and had a total salary of $135,856,866.

Team #1							
Name	NBA Team	Position	Age	Win Share	Salary	% of Games Started	All-Star
Shai Gilgeous-Alexander	OKC	PG	24	11.4	$ 33,386,850	100%	Yes
Tyrese Haliburton	IND	PG	22	7.6	$ 5,808,435	100%	Yes
Desmond Bane	MEM	SG	24	5.8	$ 3,825,083	100%	No
Immanuel Quickley	NYK	SG	23	6.7	$ 4,171,548	26%	No
Trey Murphy III	NOP	SF	22	7.6	$ 3,359,280	82%	No
Kenyon Martin	HOU	SF	22	4.9	$ 1,930,681	60%	No
Evan Mobley	CLE	PF	21	8.5	$ 8,882,640	100%	Yes
Lauri Markkanen	UTA	PF	25	8.2	$ 17,259,999	100%	Yes
Nikola Jokic	DEN	C	27	14.9	$ 47,607,350	100%	Yes
Nic Claxton	BKN	C	23	9.2	$ 9,625,000	100%	No

Table 1: Team #1 - Results of First and Second Model

Given the results from the first model, I concluded the maximum win share of a roster while having a salary under the salary cap was 84.8. The next iteration of the model was minimizing the salary while retaining the maximized amount of win share. This proved no different as the results of the model were exactly the same as the first model. As the roster construction was the same for both instances, I only counted the results as a singular team, Team #1, for the sake of simplifying the simulation process.

As mentioned, the finalized model was created to have a roster with a more realistic application to the NBA. Table 2 shows the roster results of the finalized, more realistic model. The model created a more balanced team, which will be referred to as 'Team #2'. The total win share for Team #2 was 76.2 and the roster has a salary of $130,093,997. The average age of Team #2 was 23.4 years old and for Team #1, the average age was 23.3 years old.

Team #2							
Name	NBA Team	Position	Age	Win Share	Salary	% of Games Started	All-Star
Shai Gilgeous-Alexander	OKC	PG	24	11.4	$ 33,386,850	100%	Yes
Tyus Jones	WAS	PG	26	5.7	$ 14,000,000	28%	No
Desmond Bane	MEM	SG	24	5.8	$ 3,825,083	100%	No
Immanuel Quickley	NYK	SG	23	6.7	$ 4,171,548	26%	No
Trey Murphy III	NOP	SF	22	7.6	$ 3,359,280	82%	No
Saddiq Bey	ATL	SF	23	3.9	$ 4,556,983	48%	No
Evan Mobley	CLE	PF	21	8.5	$ 8,882,640	100%	Yes
Santi Aldama	MEM	PF	22	4.6	$ 2,194,200	26%	No
Nikola Jokic	DEN	C	27	14.9	$ 47,607,350	100%	Yes
Onyeka Okongwu	ATL	C	22	7.1	$ 8,109,063	23%	No

Table 2: Team #2 - Results of Finalized Model

When analyzing the roster construction of both Team #1 and Team #2, the first takeaway is how the model selected six of the same players for both teams: Shai Gilgeous-Alexander, Desmond Bane, Immanuel Quickley, Trey Murphy III, Evan Mobley, and Nikola Jokic. The reason I deducted for the models selecting the same players is that they have high win share totals among their positions and the model identifies players that would improve the maximum win share total but not at a detriment to the salary total of the roster. Another interesting point to make is the average age of the selected players. Many of the players are under the age of 25 for both Team #1 and Team #2. The average age of Team #2, 23.4 years old, is close to the lower bound set for average age, which was 23.14. The reason for a younger roster correlates to younger players still playing on their rookie contracts, which have a much lower salary in comparison to players who have completed their rookie contracts. Players with a smaller contract make up a lower percentage of the salary cap while still providing viable win share totals to maximize the team's total win share.

4.2 Analyzing the Simulation Results

Using WhatIfSports.com, I successfully imported the data for Team #1 and Team #2, located the team for the 2015 – 2016 Golden State Warriors and 2022 – 2023 Denver Nuggets, and simulated games. The team data discussed for Team #1 and Team #2 including average age, number of all-stars, total win share, and total salary are shown in Table 3. As well, the dame statistics are given for the teams I simulated games against.

Team	Average Age	Number of All-Stars	Total Win Share	Total Salary	Cost / Win Share
Team #1	23.3	5	84.8	$ 135,856,866	$ 1,602,085.68
Team #2	23.4	3	76.2	$ 130,092,997	$ 1,707,257.18
'15-'16 Warriors	27.38	4	67.8	$ 93,248,622	$ 1,375,348.41
'22-'23 Nuggets	26.72	1	49.3	$ 159,119,037	$ 3,227,566.67

Table 3: Statistics for Team #1, Team #2, and Opponents

An example of the setup for simulating a game is shown in Figure 3. When loading in WhatIfSports, I selected the 'NBA' tab as this brought me to the webpage related to NBA simulations. I then created my own 'Dream Team' as the website calls it and selected the players identified from the model. To set the simulated game up, I selected my created team called 'OIE 559 Test #1', which was located under the 'Dream Team' tab, in Figure 3 as the away team, and by using the 'Historical' tab, set the 2015-16 Golden State Warriors as the home team. From here, I began the simulation of a singular game by clicking on the 'Play Game' button.

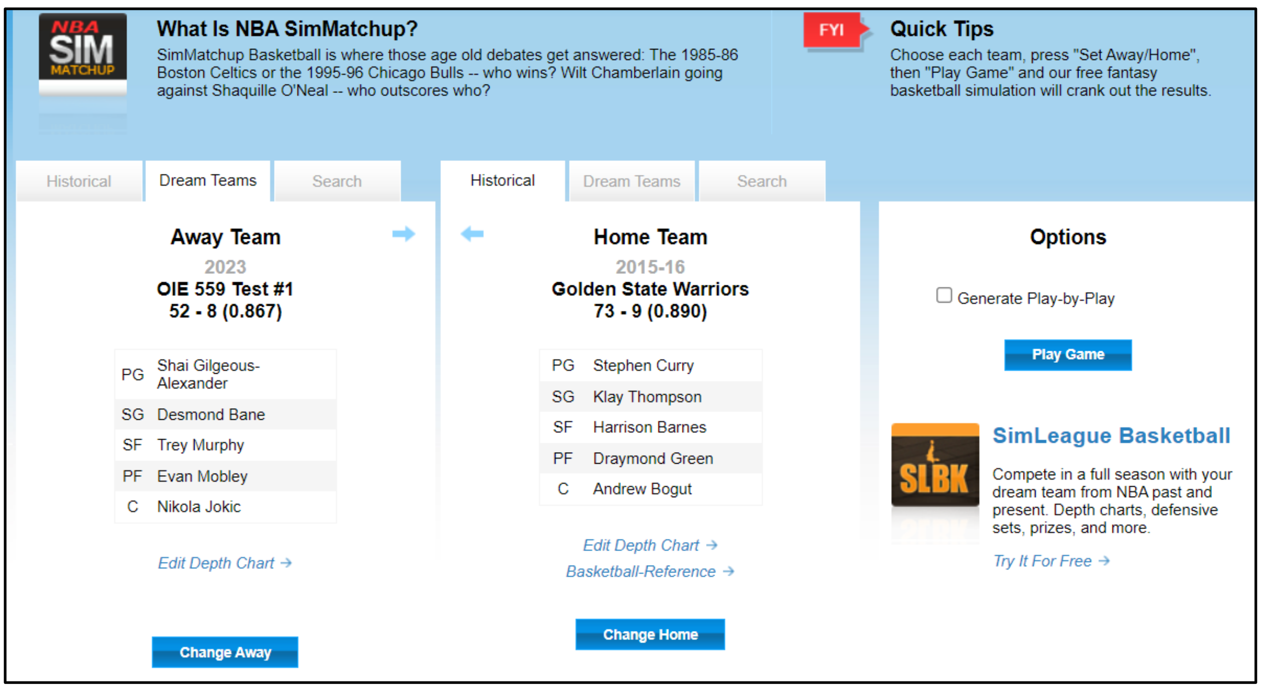

Figure 3: WhatIfSports Simulation Set Up Example

The simulation returns a final score for both teams the associated box scores for teams and individual statistics for both teams. Figure 4 illustrates the final result of the simulation. The simulation results of Figure 4 were not included in the results shown below. To keep a consistent record of the results of the sets of games, the results were logged on a game-by-game basis to Microsoft Excel.

Final - 10/11/2023	
OIE 559 Test #1	114
2015-16 Golden State Warriors	104

Player of the Game

Nikola Jokic (OIE 559 Test #1)

20 pts, 12 reb, 6 ast, 1 blk

	1	2	3	4	T
OIE 559 Test #1	24	30	28	32	114
2015-16 Golden State Warriors	17	31	26	30	104
	play-by-play	play-by-play	play-by-play	play-by-play	

Historical match-ups are awesome, but you can take it a step forward with SimLeague Basketball. Build your own dream team of historical players and then compete for prizes against other members in a full-season. Leagues simulate 2 games per day and include extensive stat tracking and gm options. There's also the X's and O's stuff -- depth charts, double-teams, temp, defensive sets and more!

Get started in your FREE exhibition season of SimLeague Basketball.

OIE 559 Test #1

Starters	MIN	FGM/FGA	3PM/3PA	FTM/FTA	OREB	REB	AST	BLK	STL	TO	PF	PTS
2022-23 Nikola Jokic	36	9/13	1/1	1/3	2	12	6	1	0	6	4	20
2022-23 Shai Gilgeous-Alexander	33	9/14	0/0	4/5	0	4	5	0	0	3	3	22
2022-23 Evan Mobley	33	5/9	0/1	1/3	2	8	2	1	1	0	1	11
2022-23 Desmond Bane	26	4/8	2/3	2/2	0	9	6	1	1	0	0	12
2022-23 Trey Murphy	26	4/7	3/5	0/0	2	5	1	1	0	0	1	11
Bench	**MIN**	**FGM/FGA**	**3PM/3PA**	**FTM/FTA**	**OREB**	**REB**	**AST**	**BLK**	**STL**	**TO**	**PF**	**PTS**
2022-23 Kenyon Martin	22	3/6	1/2	0/0	0	4	1	0	1	1	0	7
2022-23 Immanuel Quickley	22	2/7	1/3	0/0	0	2	4	0	0	0	1	5

Figure 4: Example Result of WhatIfSports Simulation

Through conducting 30-game simulations with alternating the home team and away team after 15 games, I gained insight into the performance of the optimization models and the team's expected performance on the court. Table 4 shows the results for the simulated games for Team #1 and Team #2.

	Team #1					
Opponent	W	L	Pct.	Avg. PTS/G	Avg. Opp PTS/G	PPG Diff
'15-'16 Warriors	27	3	0.900	120.47	104.30	16.17
'22-'23 Nuggets	25	5	0.833	120.13	110.20	9.93
	Team #2					
Opponent	W	L	Pct.	Avg. PTS/G	Avg. Opp PTS/G	PPG Diff
'15-'16 Warriors	23	7	0.767	113.93	103.43	10.50
'22-'23 Nuggets	18	12	0.600	115.33	108.17	7.17

Table 4: Simulation Results for Team #1 and Team #2

Team #1 had great success against all competitors as they eclipsed a total winning percentage of 0.867, which would be a top three winning percentage in NBA history ("List of National Basketball Association Seasons," 2023). In terms of average points per game, Team #1 would have placed second in the entire league for the 2022 – 2023 season in points per game only behind the Sacramento Kings, who averaged 120.7 points per game (*Teams Traditional | Stats | NBA.Com*, n.d.). For opponent's points per game, which is the total number of points a team allows, Team #1 allowed 107.25 points per game. At this total, Team #1 would have placed second in the lowest number of points allowed to an opponent, only being behind the Cleveland Cavaliers, who allowed 106.9 points per game (*Teams Opponent | Stats | NBA.Com*, n.d.). Team #1 shows their true dominance in their team's points-per-game difference, commonly referred to as Plus/Minus. This statistic is the difference between the number of points per game scored and the number of allowed points per game. Team #1 had a difference of 13.05 points per game, which would be more than double than the best team in the 2022 – 2023 season, the Boston Celtics at 6.5 points (*Teams Traditional | Stats | NBA.Com*, n.d.). The sheer dominance of Team #1 contributes to the unrealistic features of the roster construction.

To analyze the performance of Team #2, a similar comparison of statistics can be conducted. Team #2 was not as dominant as Team #1 as the model was created to better resemble an NBA team that could be assembled in today's league. Team #2 had an overall winning percentage of 0.682, which would make the team a top-three competitor during the 2022 – 2023 season (*Teams Traditional | Stats | NBA.Com*, n.d.). Team #2's points per game total was 114.63 points, which would make the team a league average offensive team (*Teams Traditional | Stats | NBA.Com*, n.d.). Team #2's strength comes on the defensive side of the ball as they only allowed their opponents 105.8 points per game, which would rank at the top of the NBA during the 2022 – 2023 season (*Teams Opponent | Stats | NBA.Com*, n.d.). With these two statistics in mind, their points per game difference is 8.83, which would total above the top performer, the Boston Celtics at 6.5 points (*Teams Traditional | Stats | NBA.Com*, n.d.). Team #2 does not have as impressive a winning percentage and offensive capability as Team #1 but does show how effective optimization is to assemble realistic NBA teams.

One concern of the analysis is that the level of competition the modeled created teams are simulating against is higher than the league average, which could result in deflated statistics. Furthermore, as the number of simulated games for each team sums to sixty games, a typical NBA season has 82 games played for each team, which could cause a less than perfect representation of the team's performance. For this reason, winning percentage, instead of a team's record, is used as the comparison statistic. With the two concerns aside, the performance of both teams is significant enough to highlight the impact optimization can have on the NBA and the decision-making processes utilized by teams when discussing player selection.

5. Closing Statements

5.1 Recommendations

For future work on this project, there are great opportunities to provide value in the NBA industry. My first recommendation would be to take the current models and use WhatIfSports to simulate an entire season. I would take both Team #1 and Team #2 and import them into separate 2022 – 2023 season simulations, removing a team that finished the year with a winning percentage in the bottom quarter of the league. I recommend simulating around three to five times to gather more data and record the results for each season. After this, I would suggest conducting analysis and drawing more conclusions based on the performance of the teams.

The next recommendation I have to improve the project is to implement more constraints to create more realistic scenarios or specific scenarios where insights can be drawn to the outcomes. I would take the output roster and conduct a similar simulation as discussed above. By creating more realistic scenarios, obtaining data for the incoming rookie class would help plan for the coming year and instead of focusing on creating the best team for a singular season, I would recommend building a model to look at the longevity of the team. For example, as the current model is focused on the 2023 – 2024 season, there is a weakness in longevity as the younger players who have strong performance based on win share are typically on their rookie deals. In a span of three to five years, these rookie deals would have expired and the team would have to make sacrifices to obtain the players while remaining under the salary cap as their financial compensation would increase.

The last recommendation I have is to look further into why players keep repeating on the roster iterations and create a model that can take the outputs of the previous iterations and assign negative weights to them, so there can be new analysis done on new rosters. This may prove to be beneficial.

5.2 Conclusion

In closing, implementing binary integer programming models is effective in constructing successful NBA rosters. By maximizing the total win share (WS) of a roster, the model outputs teams who have a high ceiling of success while creating a balanced team, which remains below the salary cap for the 2023 – 2024 NBA season and the previously determined roster constraints. By taking an iterative approach to model the problem, updates to the model were included to add constraints that would improve the realistic feasibility of creating this team in the NBA. The success of the produced rosters during simulation of the different sets of games further demonstrates the impact win share has as an effective metric in assessing a player's ability to influence winning and eventually, be a determinant in decision-making on a player's expected performance. Furthermore, the results provide evidence for integer programming being a viable tool in the sports analytics industry.

6. Work Cited

2015-16 Golden State Warriors Schedule | Basketball-Reference.com. (n.d.). Retrieved September 20, 2023, from https://www.basketball-reference.com/teams/GSW/2016_games.html

2022-23 Denver Nuggets Roster and Stats | Basketball-Reference.com. (n.d.). Retrieved September 20, 2023, from https://www.basketball-reference.com/teams/DEN/2023.html

2022-23 NBA Player Stats: Per Game | Basketball-Reference.com. (n.d.). Retrieved October 1, 2023, from https://www.basketball-reference.com/leagues/NBA_2023_per_game.html

Editor, M. C. (2018, January 16). *How NBA Analytics is Changing Basketball | Merrimack College*. Merrimack College Online. https://online.merrimack.edu/nba-analytics-changing-basketball/

History. (n.d.). NBA Careers. Retrieved October 1, 2023, from https://careers.nba.com/history/

Holt, Griffin. (n.d.). *Drafting the best baseball team: an integer programming problem.* Retrieved September 3, 2023. *https://cap.stanford.edu/profiles/cwmd?fid=301672&cwmId=10836.*

List of National Basketball Association seasons. (2023). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_National_Basketball_Association_seasons&oldid=1159887620

NBA Roster Survey: Facts to know for the 2022-23 season | NBA.com. (n.d.). Retrieved September 28, 2023, from https://www.nba.com/news/nba-roster-survey-facts-to-know-for-the-2022-23-season

NBA sets salary cap at $136 million for 2023-24 season | NBA.com. (n.d.). Retrieved September 20, 2023, from https://www.nba.com/news/nba-sets-salary-cap-at-136-million-for-2023-24-season

NBA Win Shares. (n.d.). Basketball-Reference.Com. Retrieved September 20, 2023, from https://www.basketball-reference.com/about/ws.html

Our Leagues. (n.d.). NBA Careers. Retrieved October 10, 2023, from https://careers.nba.com/our-leagues/

SimMatchup Basketball—Free NBA Matchup and Basketball Sim Games. (n.d.). Retrieved September 20, 2023, from https://www.whatifsports.com/nba/default.asp

Teams Opponent | Stats | NBA.com. (n.d.). Retrieved October 9, 2023, from https://www.nba.com/stats/teams/opponent

Teams Traditional | Stats | NBA.com. (n.d.). Retrieved October 9, 2023, from https://www.nba.com/stats/teams/traditional?dir=A&sort=PLUS_MINUS

The NBA's Adam Silver: How Analytics Is Transforming Basketball. (n.d.). *Knowledge at Wharton*. Retrieved October 10, 2023, from https://knowledge.wharton.upenn.edu/article/nbas-adam-silver-analytics-transforming-basketball/

Thompson, Green among reserves for 2018 NBA All-Star game. (n.d.). Retrieved September 28, 2023, from https://www.mercurynews.com/2018/01/23/klay-thompson-draymond-green-among-reserves-for-2018-nba-all-star-game/