Paper 1790-2014

Money Basketball: Optimizing Basketball Player Selection Using SAS®

Sabah Sadiq, Deloitte Consulting LLP, Chicago, IL; Jing Zhao, Deloitte Consulting LLP, Raleigh, NC

ABSTRACT

Over the past decade, sports analytics has seen an explosion in research and model development to calculate wins, reaching cult popularity with the release of Moneyball. Drawing inspiration from the Maximum Coverage Problem (MCP) in Operations Research, we will explore the methodology and possible solutions to a real-life “Moneyball” problem in basketball drafting.

We aim to determine a player’s worth through clustering their shots together into “sweet spots” and estimating the number of shots per game within each sweet spot. This data will then be fed into an optimization algorithm to select an optimal basketball lineup that maximizes total field goals per game along with court coverage. Finally, additional constraints will be added into the model that solve for three different scenarios.

INTRODUCTION

BUSINESS QUESTION

Imagine you are a basketball coach picking a new lineup, or a fantasy basketball aficionado drafting his/her dream team. Your end goal is to win with a strong team of players, right? While it seems a no-brainer to keep your personal loyalties out of it and keep LeBron James in the roster, there’s more to selecting a winning team. Is LeBron the “biggest bang for the buck”, when you have a constraint on budget? Is he adding new value to the other 4 players you have in mind? This paper aims to present a fresh perspective to traditional player selection that uses optimization to achieve the highest possible field goals (two or three point basket excluding free throws) per game.

MODEL OVERVIEW

This paper presents a model with a new “twist” of the classical maximal service distance in Operations Research. Selecting a location for emergency facilities (such as fire stations or ambulance dispatching stations) with the maximum service distance can allow for cost-effectiveness and faster response time. Similarly, in a basketball game, a team optimizes its chances for winning by maximizing court coverage and selecting players with a higher probability of scoring.

SAS/OR® (Operations Research) package uses threaded processing to solve linear and nonlinear programming problems efficiently. This paper builds the model with PROC OPTMODEL, a key procedure in SAS/OR®.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

MODEL INPUTS & OUTPUTS (DATA)

Raw data is collected from Basketball Geek (Basketball Geek, 2013), a public website hosted by a PhD candidate in statistics, in a play-by-play format for all 30 NBA teams during the time period of Oct 2009 to April 2010. Visual Basic for Applications (VBA) macros are created to combine the 1,215 Excel .csv files into one consolidated table and imported into SAS®. This data provides the X and Y coordinates for every shot taken, player name and points earned if the shot is made. The court size is 50x70 where (0, 0) begins at the upper left side of the court (opponent court corner near the coaching box bench). The center of the offensive hoop is located at (25, 5.25). Table 1 shows a sample of this data.

Period	Time	Player Name	Points Earned	X	Y
1	4:00	Paul Pierce	2	25	6
1	5:45	Kevin Garnett	2	27	24
1	6:12	Kendrick Perkins	2	25	6
1	6:53	Rajon Rondo	2	25	6
1	7:18	Ray Allen	2	25	6
1	10:05	Kevin Garnett	2	17	6
1	11:41	Paul Pierce	2	25	6

Table 1. Raw Data Snippet (Boston Celtics vs. Houston Rockets, Period 1)

This raw data doesn't take into account the fact that players have strong and weak spots. A player may shoot from the right side of the court often but miss the majority of their shots and vice versa. Thus, the data is analyzed through clustering shots made by each player into "Sweet Spots." K-means clustering is used to group similar X and Y coordinates together using PROC FASTCLUS. In order to expedite the optimization algorithm, the MAXCLUSTERS option is used to restrict the maximum number of clusters to 3 for each player. After determining a player's sweet spots, the average field goals are calculated for each player's sweet spot. Table 2 displays this data for three popular players.

Player	Sweet Spot	MinX	MinY	MaxX	MaxY	No. of Games	Total Field Goals	Avg. Field Goals Per Game
Kobe Bryant	1	21	4	45	24	68	686	10.0
Kobe Bryant	2	4	5	20	21	51	222	4.5
Kobe Bryant	3	11	13	35	29	52	240	4.7
LeBron James	1	5	6	23	27	46	172	3.7
LeBron James	2	16	3	43	18	74	888	12.0
LeBron James	3	24	7	46	29	38	172	4.5
Shaquille O'Neal	1	30	6	39	15	8	20	2.5
Shaquille O'Neal	2	23	5	30	16	52	434	8.3
Shaquille O'Neal	3	16	6	23	15	25	64	2.5

Table 2. Modeling Dataset Sample

The sweet spots above are displayed below visually for Kobe Bryant, LeBron James, and Shaquille O'Neal respectively. Each dot represents a clustered spot in the court where the player scores at least once. Note that the points earned per dot are not represented in the graph below.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

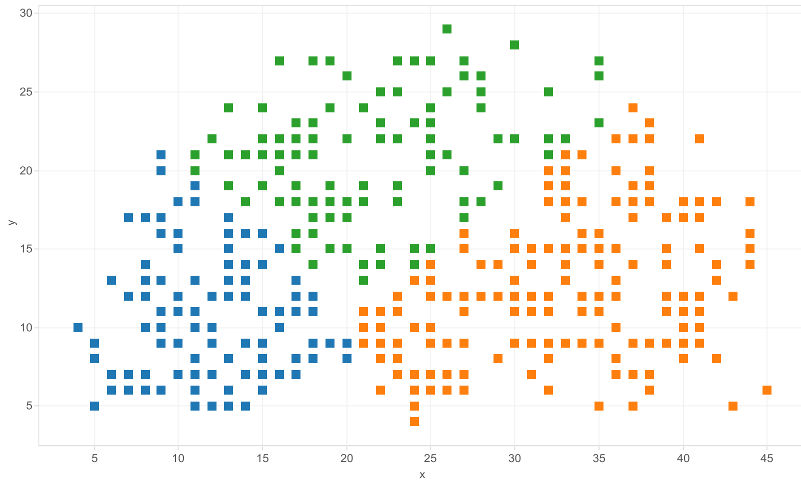

Figure 1. Kobe Bryant "Sweet Spot" Map

Kobe Bryant has the largest court coverage and 10 points per game on average for the 1st cluster (blue area).

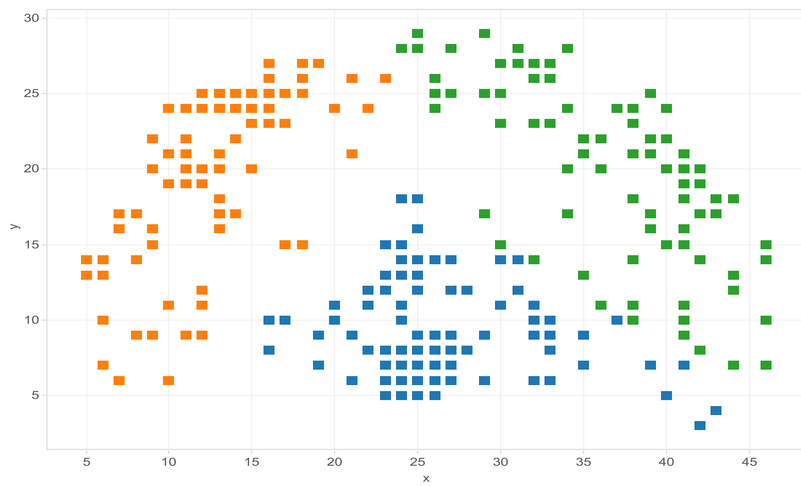

Figure 2. LeBron James "Sweet Spot" Map

LeBron James has the highest points per game in the 2nd cluster (orange area).

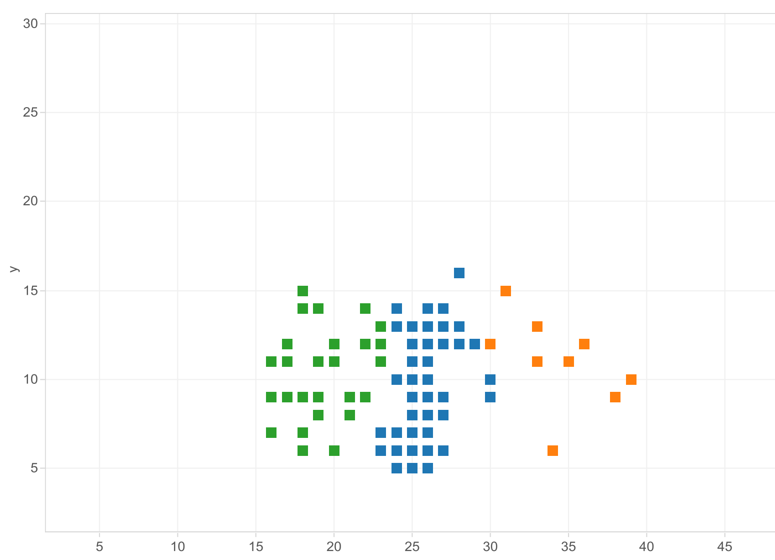

Figure 3. Shaquille O'Neal "Sweet Spot" Map

Shaquille O'Neal has slightly better coverage on one side of the court (green area) than the other. He scores fewer points per game compared to the two other players aforementioned.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

THEORY

The problem presented falls under a classical optimization problem called the Maximum Coverage Problem (MCP). MCP is a problem where you want to select at most k of the inputs to cover as much ground as possible. The traditional MCP program has been altered to maximize game field goals instead of coverage. The court coverage is instead maximized while the selected players do not have overlapping sweet spots. Additionally, the traditional MCP formulation has been altered to accommodate the sweet spot aggregation of the dataset. Although, traditional MCP formulations are NP-hard, these alternations to the formulation allow for the model to use standard linear programming rather than approximations to arrive at an optimal solution in a reasonable amount of time. Table 3 outlines the mixed integer linear program used in this problem.

$$\begin{array}{c} Maximize \\ Total\ Game \\ Field\ Goals \end{array} = \sum_{\mathrm{i}=1}^{players} \sum_{k=1}^{sweetspots} SweetSpotAssign_{i,k} * AvgFieldGoals_{i,k} \quad (1)$$

$$Subject\ to: \quad \sum_{\mathrm{i}=1}^{players} SweetSpotAssign_{i} \leq 5 \quad (2)$$

$$\sum_{\mathrm{i}=1}^{players} \sum_{k=1}^{sweetspots} SweetSpotAssign_{i,\,k} \leq 1 \quad (3)$$

$$\sum_{\mathrm{x}=1}^{players} \sum_{k=1}^{sweetspots} If\ SweetSpotAssign_{i,\,k} = 1\ then \quad \sum_{\mathrm{x=minX}}^{\mathrm{maxX}} \sum_{y=minY}^{maxY} GridPointOccupied_{x,\,y} = 1 \quad (4)$$

$$\sum_{\mathrm{x}=0}^{50} \sum_{y=0}^{70} GridPointOccupied_{x,\,y} \leq 1 \quad (5)$$

$$SweetSpotAssign_{i,k} \in (0,1),\ GridPointOccupied_{i,j} \in (0,1), \quad (6-7)$$

Table 3. Maximum Coverage Formulation

Where:
players = *i* = Number of players = 440
sweetspots = *k* = Number of sweet spots for a player = 3
x = x axis of the court ranging from 0 to 50
y = y axis of the court ranging from 0 to 70
minX = minimum x coordinate for a player's sweet spot
maxX = maximum x coordinate for a player's sweet spot
minY = minimum y coordinate for a player's sweet spot
maxY = maximum y coordinate for a player's sweet spot
$SweetSpotAssign_{i,k}$ = Binary decision variable to assign Player *i* to Sweet Spot *k*
$GridPointOccupied_{x,y}$ = Binary variable to determine if grid point (x,y) is occupied
$AvgFieldGoals_{i,k}$ = Estimated field goals for Player *i* to Sweet Spot *k*

Copyright© 2014 Deloitte Development LLC. All rights reserved.

In the above formulation, Equation (1) is the objective function and formulates the total game field goals by iterating through the chosen players and their estimated points at their assigned sweet spot. It should be noted that only two and three pointers are included in this calculation with free throws on top of counted baskets not included. Equation (2) provides that only five players are selected. Equation (3) provides that each player is only assigned to one of their sweet spots. Equation (4) provides that a grid point is deemed "occupied" if a player has been assigned to a sweet spot that covers that grid point. This equation flattens the circular sweet spots into rectangular form in order to simplify the mathematical optimization process. Equation (5) provides that players are not assigned to overlapping grid points. This constraint can be somewhat limiting as even a slight overlap, which may not be that important, could force a valuable player out of the lineup. As a future enhancement, some attention should be given into possibly relaxing this constraint to allow for slight overlaps.

Although this linear program answers important questions about optimal player lineup, there are certain business questions that may remain unanswered in this "unconstrained" model. To this purpose, we have created different scenarios that introduce additional constraints to satisfy certain business requirements. We will produce possible solutions for all three scenarios. It should be noted that these scenarios all revolve around a purely offensive strategy with no consideration to player position or defensive skills. This is an obvious simplification of truth but is meant only to provide insights into interesting offensive player combinations, not to pick the best overarching team.

SCENARIO 1: "UNCONSTRAINED" MODEL

In this scenario, the only changes made to the traditional MCP linear program are to convert grid point assignment to sweet spot assignments and convert the objective function to maximize points. Thus, the only question that is answered is given a certain set of players, what is the optimal subset of players that a team should choose to maximize the estimated game points – regardless of any business constraints.

SCENARIO 2: "SELECT PLAYER X" MODEL

In this scenario, we address the question of what a team should do if they know for a fact they want a particular player in the lineup regardless of what the model determines. In this instance, we introduce an additional constraint that forces LeBron James in the model and determines what remaining players should be chosen to create an optimal team.

$$SweetSpotAssign_{i=Lebron\ James,k=1} = 1 \qquad (8)$$

SCENARIO 3: "BUDGET" MODEL

Finally, we address the question of what a team should do if they have a certain budget. Player salaries are downloaded from ESPN and incorporated into the dataset. An additional constraint is then added so that the total cost of the starting lineup does not exceed $30M.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

$$\sum_{i=1}^{players} \sum_{k=1}^{sweetspots} Salary_i * SweetSpotAssign_{i,k} \leq \$30{,}000{,}000 \quad (9)$$

Where:
$Salary_i$ = Cost to employ Player *i*

RESULTS

SCENARIO 1 RESULTS ("UNCONSTRAINED")

The unconstrained model chose Brian Skinner, Dwayne Wade, Gilbert Arenas, Mike Wilks and Roko Ukic as the "dream team." These players optimized total field goals at 37 points per game. During the 2009-2010 seasons, the average field goals per game are 37.69, placing our "dream team" near the average. However, the field goal estimation assumes that the player only shoots within the sweet spot that they have been assigned to. In reality, the chosen players may in fact shoot anywhere which could drastically increase the field goals per game. Figure 4 visualizes the sweet spots for the players that are chosen in this scenario. The x and y coordinates follow the same structure as the data provided with (0, 0) being the corner of the opponents court. The gray curve shows the percent of total points earned at each x for the five players. For instance, the peak value of the curve is 51% at x=25, which indicates that half of the total points are earned from the middle section of the court. From the curve, we can also see that more points are earned from the right side of the court (x>=25) than the left.

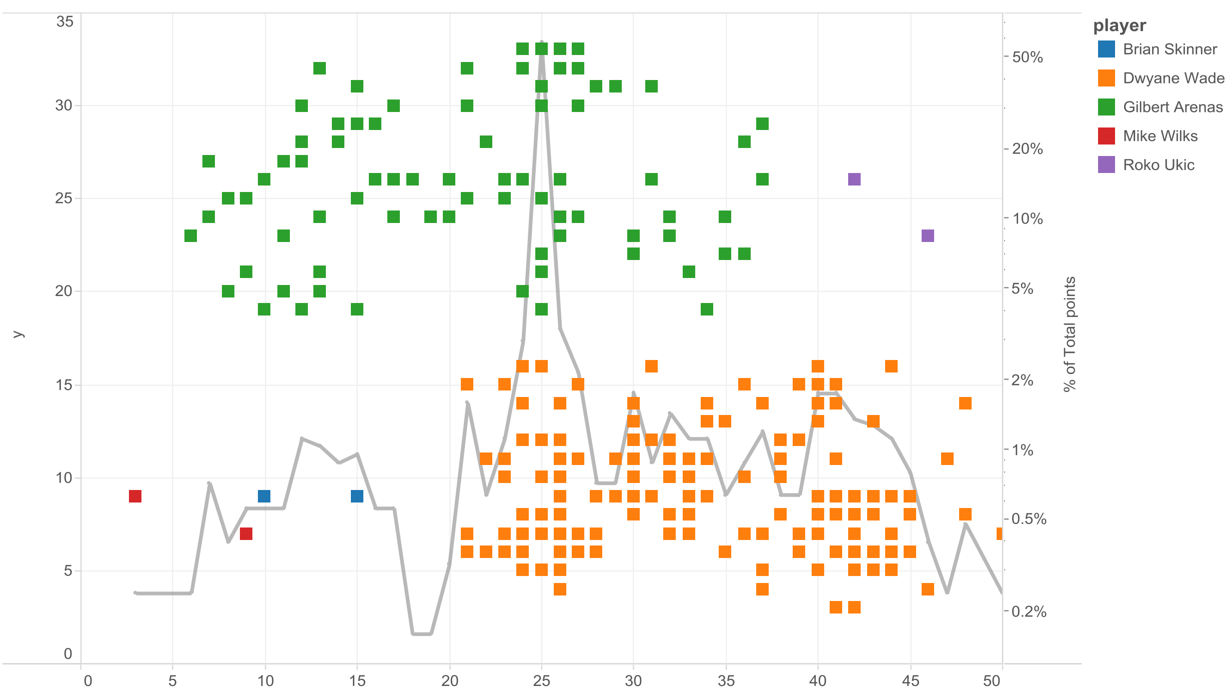

Figure 4. Scenario 1 Player Assignment

Copyright© 2014 Deloitte Development LLC. All rights reserved.

SCENARIO 2 RESULTS (“SELECT PLAYER X”)

Although Scenario 1 answers important questions regarding the offensive value of certain players it does not take into account a team owner’s preference or “instinct” to select a certain player in the lineup. To this effect, Scenario 2 forces LeBron James into the model and looks to see which players will complement the court based on LeBron’s assigned sweet spot. In this scenario, Antoine Wright, Corey Brewer, Kyle Weaver, and Marcus Williams are chosen as the other selected players. This team is estimated to accumulate 31 field goals per game. Figure 5 displays the player assignment for this scenario.

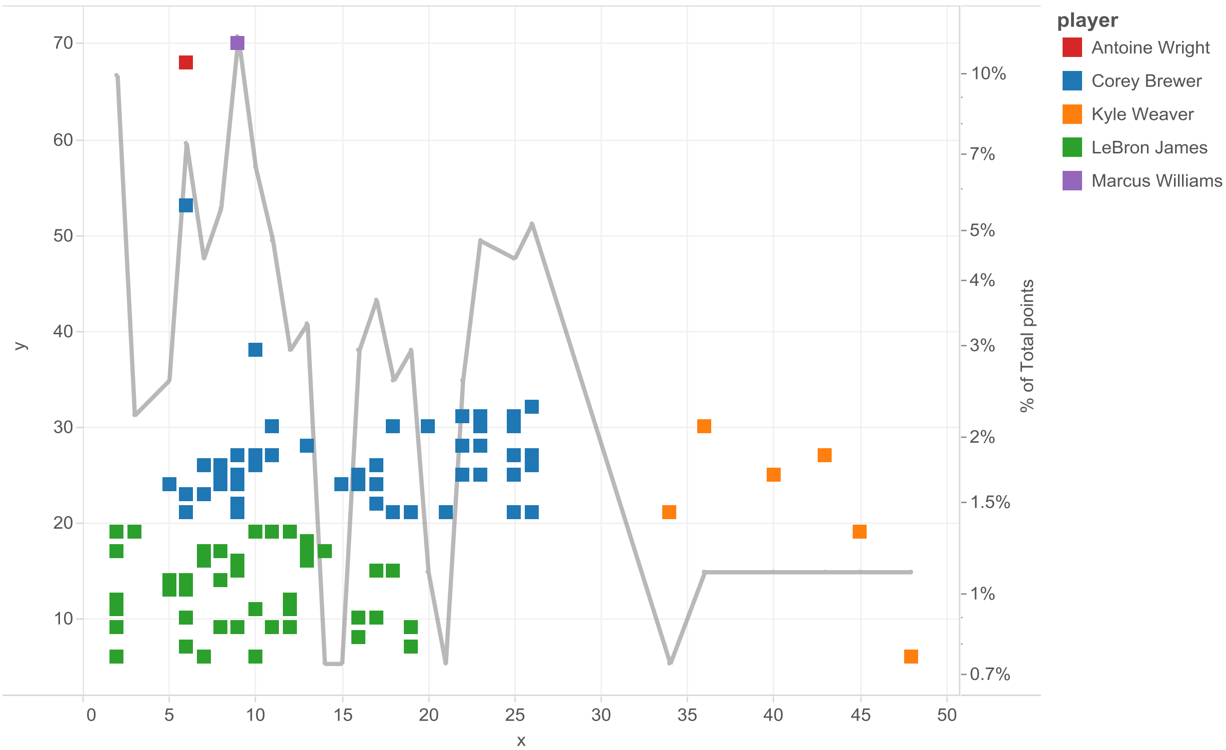

Figure 5. Scenario 2 Player Assignment

In Scenario 1, Dwayne Wade is the high shooter at 13.2 field goals in his sweet spot. In Scenario 2, LeBron is the high shooter with estimated field goals of 12.5. Coverage-wise, both players cover the same size area with no overlap; LeBron favors the right while Wade favors the left. Although they are both strong players and do not overlap sections, LeBron isn’t chosen in Scenario 1. Instead, players with lower field goals (Mike Wilks and Brian Skinner) are chosen to cover LeBron’s sweet spot. Although there could be various reasons as to why LeBron isn’t chosen, one possible explanation is that Gilbert Arenas section slightly overlapped with LeBron’s section. These results indicate that relaxing the no grid point overlap constraint could result in higher total field goals. What if a slight overlap is allowed and both LeBron and Gilbert could be chosen in the model? However, if only one can be chosen budget would certainly be a consideration which is explored in Scenario 3.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

SCENARIO 3 RESULTS ("BUDGET")

With a $30M starting lineup cap, Dwayne Wade, Andrea Bargnani, Brian Skinner and Mike Wilks are chosen. This team had a total of 29 field goals and salary of $23,442,115. Compared to the optimal team in Scenario 1 which had 37 field goals and $34,022,808 salary, this is roughly a $10M drop in salary with a 9 point drop in field goals. It is also interesting to note that with a salary cap, only four players are chosen as no additional player could add value to the team while remaining in budget. Figure 6 shows the players selected for this scenario.

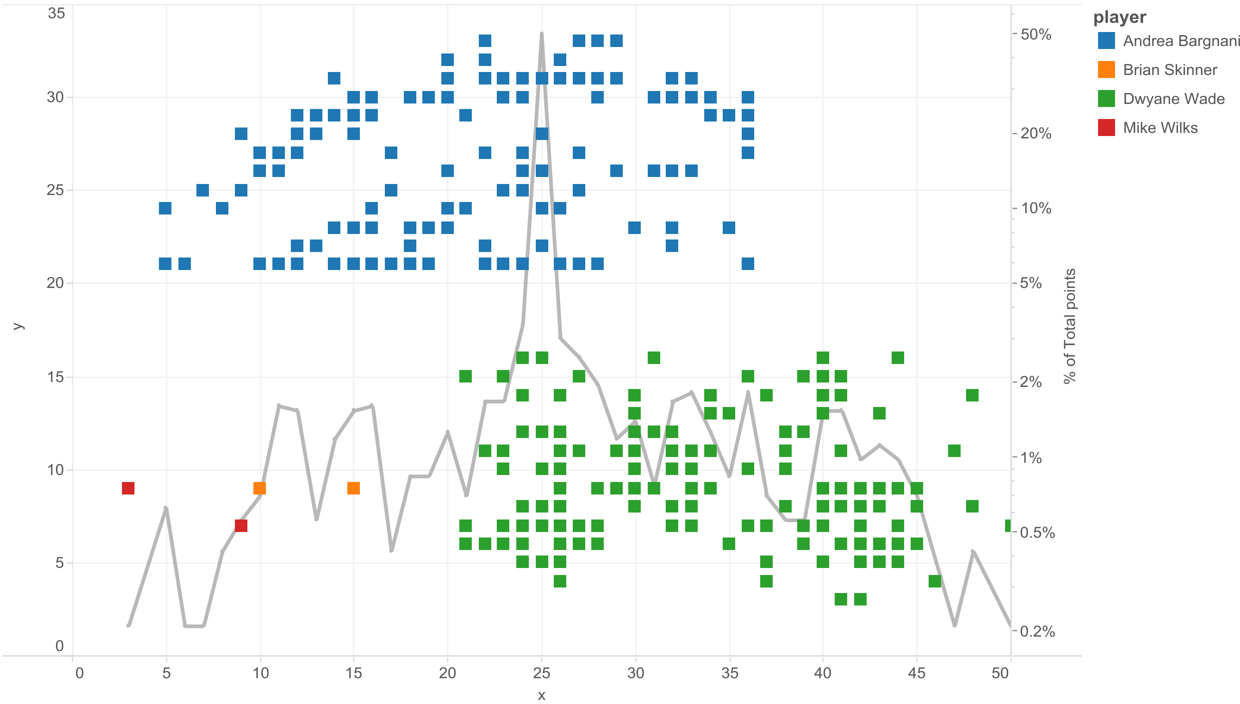

Figure 6. Scenario 3 Player Assignment

CONCLUSION

This paper briefly reviewed the Maximum Coverage Problem (MCP) in Operations Research and its application to select an "ultimate dream team lineup." The optimization scenarios explored in this paper examine different dream teams in an attempt to maximize field goals. However, a team's effectiveness involves much more than just field goals such as free throws, passing/rebounding and general team dynamics. By exploring different sports optimization scenarios, the authors hope to provide the reader with new angles for sports analytics and fantasy basketball insights.

Copyright© 2014 Deloitte Development LLC. All rights reserved.

REFERENCES

Basketball Geek. "NBA 2009-2010 Season Play by Play Data.". June 2013. Available at http://www.basketballgeek.com/downloads/2009-2010/

Church, R. and Velle, C. R. 1974. "The Maximal Covering Location Problem." Papers in Regional Science, 32: 101–118. doi: 10.1111/j.1435-5597.1974.tb00902.x

ESPN. "Field Goal Percentage Leaders 2009-2010 Season." December 17, 2013. Available at

http://espn.go.com/nba/statistics/team/_/stat/offense-per-game/sort/fieldGoalPct/year/2010

ESPN. "NBA Player Salaries - 2009-2010." December, 2013. Available at http://espn.go.com/nba/salaries/_/year/2010

SAS Publishing. 2010. "SAS/OR® 9.22 User's Guide: Mathematical Programming." Cary, NC: SAS Institute

ACKNOWLEDGEMENTS

Thanks to John Lucker, Principal at Deloitte Consulting LLP, John Houston, Principal at Deloitte Consulting LLP, David Steier, Director at Deloitte Consulting LLP, Steve Berman, Specialist Leader at Deloitte Consulting LLP, and Puli Cheng, MBA Candidate at New York University Stern School of Business, for their valuable contributions to this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Sabah Sadiq
Enterprise: Deloitte Consulting, LLP
Address: 111 S. Wacker
City, State ZIP: Chicago, IL 60606
Work Phone: 847-504-9038
Fax: 866-770-0956
E-mail: sasadiq@deloitte.com

Name: Jing Zhao
Enterprise: Deloitte Consulting, LLP
Address: 150 Fayetteville St #1000
City, State ZIP: Raleigh, NC 27601
Work Phone: 919-645-6253
Fax: 919-833-3276
E-mail: jingzhao2@deloitte.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.