

Honor Code: I affirm I have adhered to the honor code in this assignment.

## **Nguyen Tran - NBA Daily Fantasy Lineup: Optimizing Your Win Chance**

### **Abstract**

With the love for basketball, NBA is one of my favourite things to watch during free time. When considering a final project for the optimization class, I immediately considering fantasy basketball. We can combine the beauty of mathematical optimization with the NBA player data to participate in the NBA Daily Fantasy Lineup. In order to feel more 'thrill' and have better game analytics, I have register an account in DraftKings, collect a data from it to build and compete in real-money contests, hopefully can get something out of this model.

### **Rules**

- You have \$50000 to build a lineup consisting of 8 players, and must include players from at least 2 different NBA games.
- You have to choose players from a variety of positions:
  - One Point Guard (PG)
  - One Shooting Guard (SG)
  - One Small Forward (SF)
  - One Power Forward (PF)
  - One Center (C)
  - One Guard (PG or SG)
  - One Forward (SF or PF)
  - One Util (any position)

Note: One Player can play at multiple positions

- Points are awarded as follow:

Action	Points
Point	+1 Pt
Made 3pt Shot	+0.5 Pts
Rebound	+1.25 Pts
Assist	+1.5 Pts
Steal	+2 Pts
Block	+2 Pts
Turnover	-0.5 Pts
Double-Double	+1.5 Pts
Triple-Double	+3 Pts

The double-double and triple double attributes can be awarded only once per player.

## Variables

Let:

$$x_{i,p} = \begin{cases} 1 & \text{if player } i \text{ is chosen for position } p \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{i,p} = \begin{cases} 1 & \text{if player } i \text{ is eligible for position } p, \\ 0 & \text{otherwise.} \end{cases}$$

$$t_{i,t} = \begin{cases} 1 & \text{if player } i \text{ is in team } t, \\ 0 & \text{otherwise.} \end{cases}$$

$s_i$  is the salary of player  $i$

$y_i$  is the average fantasy point per game of player  $i$

## IP Formulation

Maximize

$$\sum_{i=1}^n \sum_{p=1}^8 x_{i,p} \cdot y_i$$

Subject to:

$$\sum_{i=1}^n \sum_{p=1}^8 x_{i,p} = 8 \quad (\text{Roster Constraint})$$

$$\sum_{p=1}^8 x_{i,p} \leq 1 \quad \forall i \quad (\text{Player Constraint})$$

$$\sum_{i=1}^n x_{i,p} = 1 \quad \forall p \quad (\text{Position Constraint})$$

$$x_{i,p} \leq p_{i,p} \quad \forall i, p \quad (\text{Player Specific Position Constraint})$$

$$\sum_{i=1}^n \sum_{p=1}^8 x_{i,p} \cdot s_i \leq 50000 \quad (\text{Salary Constraint})$$

$$\sum_{i=1}^n t_{i,t} < 8 \quad \forall t \quad (\text{Team Constraint})$$

After some careful considerations, I decide to create another model that accounts for minutes per game of players (I expect player that play more minutes has higher potential of scoring more points), and maximize the average points to minutes ratio.

Let:

$m_i$  is the average minutes per game of player  $i$ , and we rescale  $m_i$  to avoid low-minute players:

$$m_i \sim \{0, 1\} + 1$$

The new objective function is:

Maximize

$$\sum_{i=1}^n \sum_{p=1}^8 \frac{y_i}{m_i} x_{i,p}$$

Subject to the same constraint as above.

## Data

For data, I used:

- DraftKings Salaries and Expected Fantasy PPG dataset daily on the websites
- NBA all players dataset from <https://www.nbastuffer.com/nba-stats/player/> . I update the player's data every week.

## Implementation

Using the gurobi python package, I was able to implement this linear optimization problem and output 8 players for 8 positions. Code are available here:

[https://github.com/ntran4869/linopt\\_nba/blob/master/fantasy\\_optimizer.py](https://github.com/ntran4869/linopt_nba/blob/master/fantasy_optimizer.py)

Note: In the model implementation part, I did not take account for the team constraint, since it overcomplicates the model and at the same time, the chance of choosing all players from the same team that maximizes points is slim to none.

## Problems

Some problems that I have faced while working on this project:

- Some players are marked out from the game, but still included in the dataset in DraftKings, therefore I have to manually edit them.
- Some players are marked as 'Questionable' being chosen in the roster, and they are marked 'Out' right before the match, therefore I cannot adjust the roster in time.
- There are some 'outliers' in the data, especially for players that has \$4000 or lower. Their expected Fantasy Points are too high for their values, since their stats are unreliable (some of them only play on unimportant minutes, or they was having one or two good game thus increasing their FP). My models use to choose a lot of players that is good on paper, and then scoring 0 points since they do not even appear on court.

## Results

I have participated in the competitions until Dec 14th, engaged in 36 contests, winning for a total of 7 times. The best prize I have won is \$3, and the highest position is top 4%. Here are some of the insights from these models Pts and Points/Min.

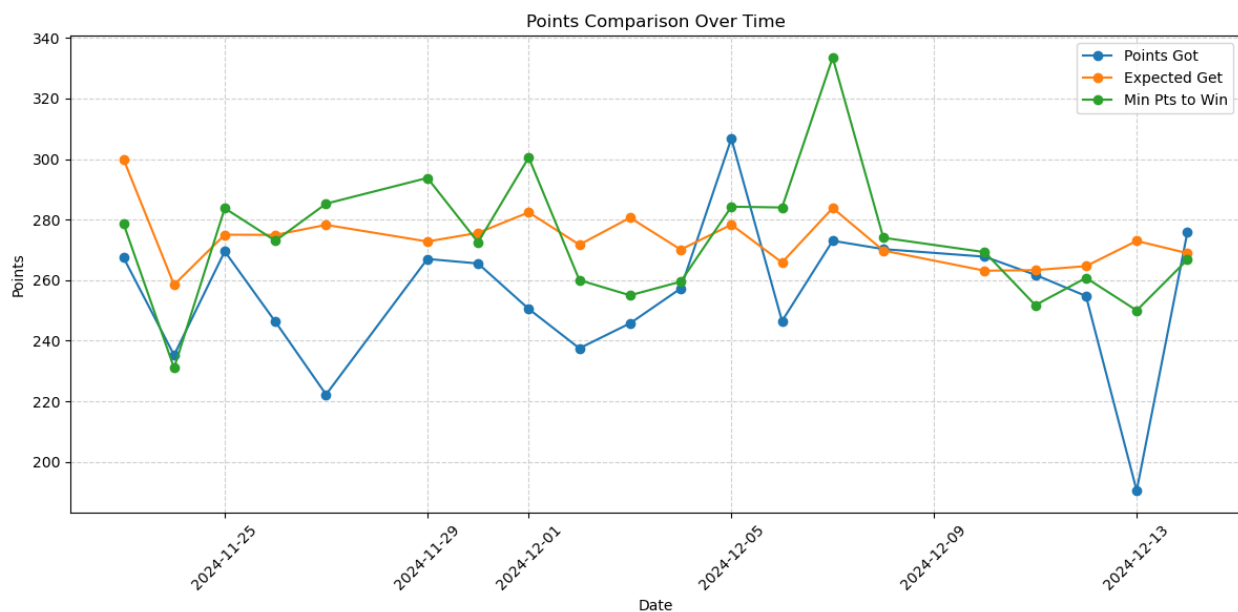


Figure 1: Points and Expected Points from Model Pts

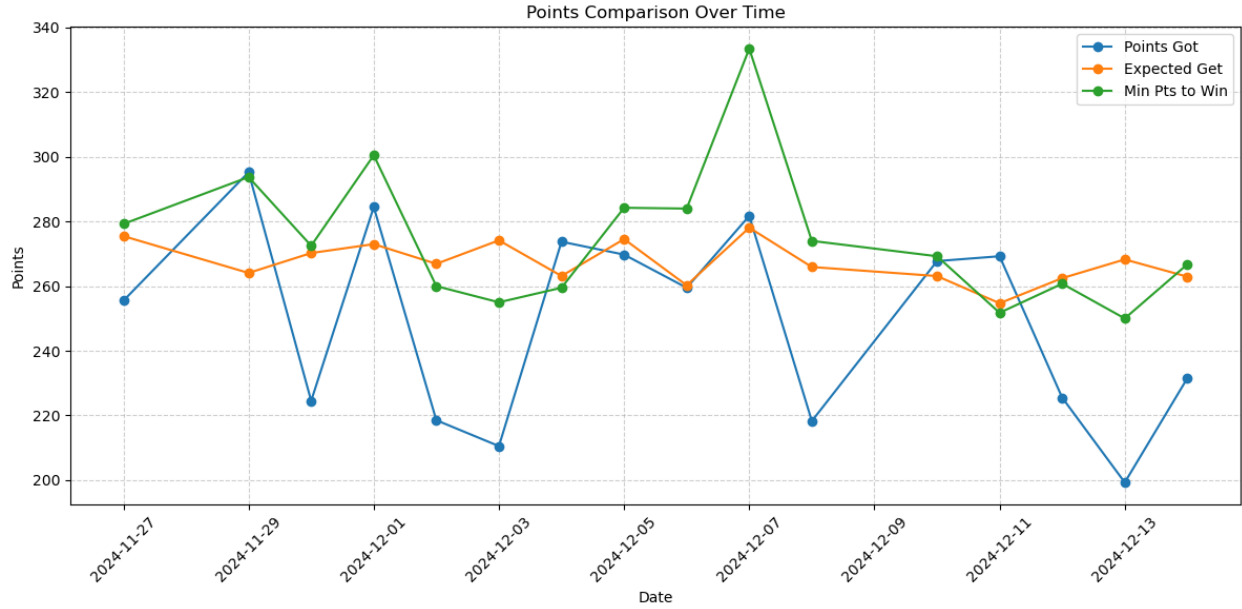


Figure 2: Points and Expected Points from Model Pts/Min

- As we can see in the graph, most of the time, most of the time, the expected points got is smaller than minimum points to win, which suggests that our model cannot take account for the variance in the player's performances. We are missing about around 20 points if we get the exact expected points in order to win.
- We can also see the trend of our real points and minimum points to win are pretty similar. This can be explained by we have chosen relatively best players there in the pool, which has high pick-rate among contestants, therefore if those players perform well, they will raise the point cap. A remarkable example is on 7th Dec, Nikola Jokic scored 90.5 points, which raises the minimum points to win up to 335.

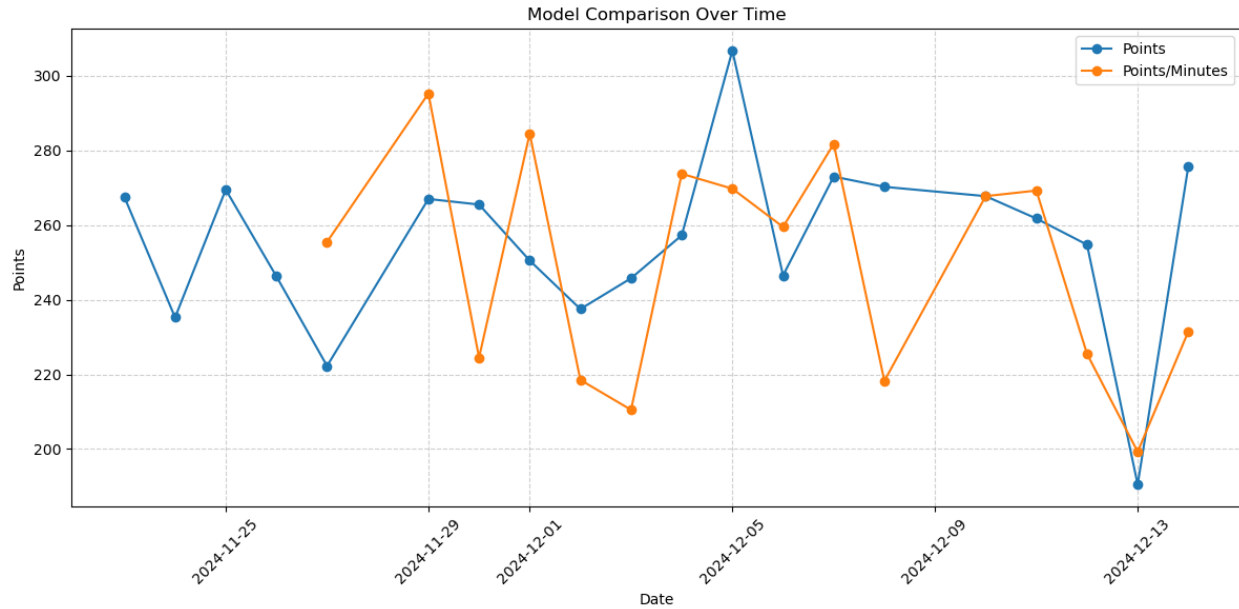


Figure 3: Points get of the two models

- The model Points performs much more consistently, the amount of points got is usually around 260, while the model Points/Minutes varies more. This is due to the problem I stated above, some of the players chosen for the second model are only good on paper, they will not perform as they should.
- From this short intervals, it is hard to say which model perform better, but my intuitions will go with the points/minutes model, with appropriate model tuning, they will generate some hidden gems players that can score a lot with low salaries.

## Takeaways and Future Work

- This is one of a few projects that I learned a lot but at the same time enjoy doing it a lot. I had to use critical thinking and analysis skills, along with losing some nerves while watching my players perform. All and all, I lost some cash (\$20), but I the knowledge I got from this worths more than it.
- Things I want to do in the future is to improve the models by getting more accurate data, account for player expected points intervals, and also do something to automate the process.