

BANA 7020-001 Fall 2018
Optimization Models
Fantasy sports League
Team Selection Optimization

By

Syed Imad Husain

M12958531

husainsd@mail.uc.edu

Abstract

This project deals with the implementation of Exact optimization for a Mixed integer problem for Fantasy sports which is a multibillion-dollar industry that gathers players' data from around the world to build virtual teams. These teams are ranked according to the real score achieved by the players in the team. Contestants compete for money or other prizes by building the best teams. In this project we deal with creating virtual Football teams for contestants. The Model is computationally solved in Xpress-IVE

Contents

The Input – Dataset	3
Data Manipulation	3
Final Data-set	4
Importing Data in Xpress IVE	4
Mixed Integer Linear Model	4
Part 1 - Model 1	4
Data and Parameters	4
Decision variable	5
Objective Function	5
Constraints	5
Part 2 – Model 2	5
Additional Data and Parameters	5
New Decision variable	5
Objective Function	6
Rewriting Constraints	6
Additional Diversity Criteria -Model 3	7
New dataset	7
Additional Data and Parameters	7
Additional Diversity Constraint	7
Conclusions and Outputs	8
Model – 1	8
Model - 2	8
Model - 3	9
Appendix	9

The Input – Dataset

To build an optimized team, the first thing we need is a data-set. The dataset provided consists of information on 100 players. Below is a summary view-

Columns	Description
Names	Player Names of the format “First Name Last Name Salutation” E.g. “Odell Beckham Jr.”
Position	Abbreviated Player positions with the following legend <ul style="list-style-type: none">• QB - Quarterback• RB - Running Back• WR = Wide Receiver• TE - Tight End• DST - Defense
Salary	Player Salary (\$). Randomly generated values between 5000-14000
Projected Score	Player Score. Randomly generated values between 1-25

Data Manipulation

For the data to be fully utilized for computation in Xpress-MP, following changes have been made. All changes were implemented using MS-Excel

- **Names** – Xpress IVE detects spaces as a new value. Since the Names column has spaces, we replace each space with a ‘_’. Also, any trailing spaces are removed for consistency purposes. This change does not change the fabric of data in any way.
- **Salary** – Generated using the following code and rounded off to the nearest integer
 $\text{ROUND}((\text{RAND}())*(14000-5000)+5000,0)$
- **Projected Score** – has also been generated in a similar fashion. However, since the range for Scores is small, we have rounded the resultant values to two decimal places for more diversity in the values
 $\text{ROUND}((\text{RAND}())*(25-1)+1,2)$
- **Position** – has been encoded as a set of five indicator fields based on the following legend

Index Position	1	2	3	4	5
DST	1	0	0	0	0
QB	0	1	0	0	0
RB	0	0	1	0	0
TE	0	0	0	1	0
WR	0	0	0	0	1

Final Data-set

Below is the screenshot of how the final data-set looks like. Each row indicates details of a single player

Name	Position	Salary	ProjectedScore	DST	QB	RB	TE	WR
Le'Veon_Bell	RB	11353	15.09	0	0	1	0	0
David_Johnson	RB	5992	9.2	0	0	1	0	0
Antonio_Brown	WR	11900	15.71	0	0	0	0	1
Alvin_Kamara	RB	5674	1.67	0	0	1	0	0
DeAndre_Hopkins	WR	5220	4.01	0	0	0	0	1
Michael_Thomas	WR	7082	22.64	0	0	0	0	1

Importing Data in Xpress IVE

As observed above, our final data-set looks like a named-matrix. It is now ready for importing in Xpress-IVE. However, to simplify the code, we break the final dataset into .txt files in the following manner –

- **names.txt** – Player Names
- **position.txt** – Player Positions
- **sal.txt** – Player Salaries
- **core.txt** – Player Projected Scores
- **matrix.txt** – Player Position matrix which contains the 5x100 values. For each player, one of the five columns will have a '1' and rest '0' to indicate at which position does the player play

Mixed Integer Linear Model

Based on the problem description we have two Linear models which are discussed below

Part 1 - Model 1

The problem to be solved is selecting the entry with maximum projected score as a linear integer program with the defined constraints

Data and Parameters

- Index i for each player with range of value $\in \{1,2,3....100\}$
- Index j for player positions with range of values $\in \{1,2,3,4,5\}$
- Set of Players P representing each player with index i
- Set for Positions W representing player position with index j
- Set of Names **Names** representing player names with index i initialized by '**names.txt**'
- Set of Positions **Positions** representing player positions with index i initialized by '**position.txt**'
- Set of Salaries **Sal** representing player salaries with index i initialized by '**sal.txt**'
- Set of Scores **Score** representing player scores with index i initialized by '**score.txt**'
- Set of Position **Matrix** Matrix representing player positions with index i, j initialized by '**matrix.txt**'

Decision variable

Binary variable X_i for each $i \in P$ such that it is

$$\begin{cases} 1 & , \text{ if player } i \text{ is selected for the team} \\ 0 & , \text{ if player } i \text{ is not selected for the team} \end{cases}$$

Objective Function

Objective is to maximize the projected score of the resultant team

$$\text{Maximize } Z = \sum_{i \in P} (X_i * \text{Score}_i)$$

Constraints

- The team should have six Players

$$\sum_{i \in P} X_i = 6$$

- The Salary of the resultant team can be 50,000 \$ at most

$$\sum_{i \in P} (X_i * \text{Sal}_i) \leq 50000$$

- There should be at least 1 player for each position

$$\sum_{i \in P} (X_i * \text{Matrix}_{i,j}) \geq 1 \text{ for each } j \in W$$

- To limit QB & DST positions to only 1 player and remaining positions can have 2 players

$$\sum_{i \in P} (X_i * \text{Matrix}_{i,j}) \leq 1 \text{ for each } j \in \text{subset of } W : \{1,2\}$$

Part 2 – Model 2

In the second part of the problem statement we have to select two teams from the pool of players. Following are the additions made to the Model in order to incorporate the requirements for the second part

Additional Data and Parameters

- Index k for each team with values $\in \{1,2\}$
- Set of Teams T representing teams with index k
- D is the number of maximum same players both team can have. It is initialized as 5

New Decision variable

- Binary variable $X_{i,k}$ for each $i \in P$ & $k \in T$ such that it is

$$\begin{cases} 1 & , \text{ if player } i \text{ is selected for team } k \\ 0 & , \text{ if player } i \text{ is not selected for team } k \end{cases}$$

- Binary variable B_i for each $i \in P$ such that it is

$$\begin{cases} 1 & , \text{ if player } i \text{ is selected for both teams} \\ 0 & , \text{ if player } i \text{ is not selected for both teams} \end{cases}$$

Objective Function

Objective is to maximize the projected score for both team

$$\text{Maximize } Z = \sum_{i \in P, k \in T} (X_{i,k} * \text{Score}_i)$$

Rewriting Constraints

- Each team should have six Players

$$\sum_{i \in P} X_{i,k} = 6 \text{ for each } k \in T$$

- The Salary of each resultant team can be 50,000 \$ at most

$$\sum_{i \in P} (X_{i,k} * \text{Sal}_i) \leq 50000 \text{ for each } k \in T$$

- There should be at least 1 player for each position

$$\sum_{i \in P} (X_{i,k} * \text{Matrix}_{i,j}) \geq 1 \text{ for each } j \in W \text{ \& } k \in T$$

- To limit QB & DST positions to only 1 player and remaining positions can have 2 players

$$\sum_{i \in P} (X_{i,k} * \text{Matrix}_{i,j}) \leq 1 \text{ for each } j \in W \text{ subset : } \{1,2\} \text{ \& } k \in T$$

- The sixth player is a flexible player that can be either a RB, WR, or TE

$$\sum_{k \in T} X_{i,k} \leq 1 + B_i \text{ for each } i \in P$$

- We can have maximum 5 players common in both teams so that both teams are not the same

$$\sum_{i \in P} B_i \leq 5$$

Additional Diversity Criteria -Model 3

As a Business Analysts we are interested in generating business for the football league. With sporting events like these, money comes in from the number of people viewing these events. The more people, the more sponsors, the more advertising the greater the business impact the event has on the community. To drive more people to view the football game in our fantasy competition we introduced a criteria of player popularity – an index from 1 to 10 using the formula in Excel

RANDBETWEEN(1,10)

New dataset

Under the assumption that more popular players will draw a larger audience. This is how the new dataset looks like

Name	Position	Salary	ProjectedScore	DST	QB	RB	TE	WR	Popularity
Le'Veon_Bell	RB	6425	7.74	0	0	1	0	0	2
David_Johnson	RB	10553	14.81	0	0	1	0	0	8
Antonio_Brown	WR	10741	16.91	0	0	0	0	1	5
Alvin_Kamara	RB	8346	20.69	0	0	1	0	0	10

Additional Data and Parameters

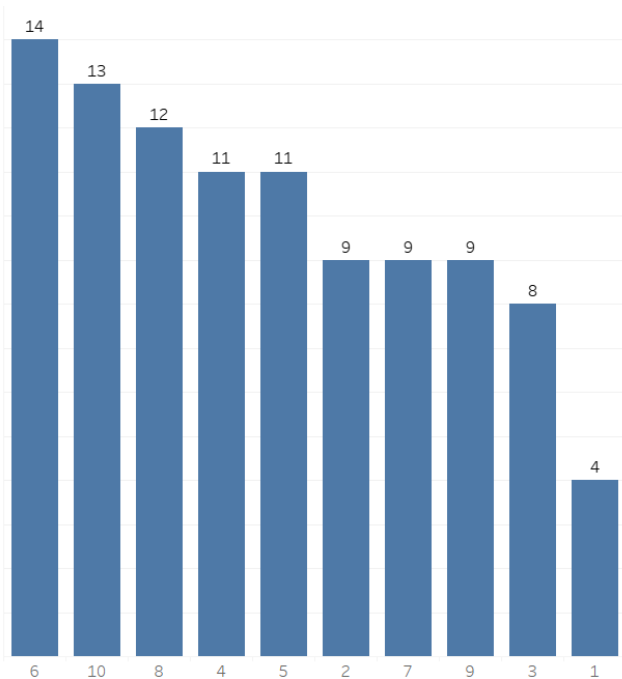
Set of Popularity **Popul** representing player popularity with index i initialized by '**popul.txt**'

Additional Diversity Constraint

We state that each team's popularity must be at least 50 points. The reason to set the diversity constraint to a higher value is because the Median of popularity is 6. Hence the average popularity will come out to be 36.

$$\sum_{i \in P} (X_{i,k} * popul_i) \geq 50$$

for each $k \in T$



Conclusions and Outputs

Model – 1

The structure of selected team is

Details	Index	Name	Pos	Salary	Score
Player 1	14	Adam Thielen	WR	9878	23.76
Player 2	57	Blaine Gabbert	QB	6348	23.63
Player 3	63	Royce Freeman	RB	5878	24.21
Player 4	97	Steelers	DST	6869	19.16
Player 5	98	Mark Walton	RB	9680	23.96
Player 6	100	David Njoku	TE	10324	19.88
Total Salary	48977				
Total Projected Score	134.6				

Model - 2

The structure of selected teams are (common players are highlighted in green)

Details	Team 1				
	Index	Name	Pos	Salary	Score
Player 1	14	Adam Thielen	WR	9878	23.76
Player 2	57	Blaine Gabbert	QB	6348	23.63
Player 3	63	Royce Freeman	RB	5878	24.21
Player 4	97	Steelers	DST	6869	19.16
Player 5	98	Mark Walton	RB	9680	23.96
Player 6	100	David Njoku	TE	10324	19.88
Total Salary	48977				
Total Projected Score	134.6				

Team 2				
Index	Name	Pos	Salary	Score
14	Adam Thielen	WR	9878	23.76
15	Kareem Hunt	RB	10907	24.16
37	Evan Engram	TE	9049	18.94
57	Blaine Gabbert	QB	6348	23.63
63	Royce Freeman	RB	5878	24.21
97	Steelers	DST	6869	19.16
48929				
133.86				

Note - If you change it to a different value for variable D it will result in a different constraint for having same players in both teams. During this iteration, it is set to 5. If we set it to 0 there will no repeated players between the two teams

Model - 3

Details	Team 1				
	Index	Name	Pos	Salary	Score
Player 1	28	Jimmy Garoppolo	QB	8683	24.7
Player 2	37	Evan Engram	TE	9049	18.94
Player 3	64	Chris Carson	RB	8613	16.96
Player 4	72	Dante Pettis	WR	8606	19.91
Player 5	80	Curtis Samuel	WR	5386	16.2
Player 6	97	Steelers	DST	6869	19.16
Total Salary	47206				
Total Projected Score	115.87				
Total Popularity	51				

Team 2				
Index	Name	Pos	Salary	Score
28	Jimmy Garoppolo	QB	8683	24.7
37	Evan Engram	TE	9049	18.94
38	Kenny Stills	WR	10544	20.28
64	Chris Carson	RB	8613	16.96
80	Curtis Samuel	WR	5386	16.2
97	Steelers	DST	6869	19.16
49144				
116.24				
50				

Note – As you can observe, the performance has come down because we introduced an additional constraint of player popularities

Appendix

- Three Xpress IVE (.mos) Files are attached to the submission on Blackboard
 - Final Project – Part A (Model 1)
 - Final Project – Part B (Model 2)
 - Final Project – Part B + Diversity Criteria (Model 3)
- One Excel File is attached PlayersData.xls – containing the final dataset
- Six Text Files are attached – which is the final Dataset split into .txt for input to Xpress-IVE
 - names.txt
 - position.txt
 - sal.txt
 - score.txt
 - matrix.txt
 - popul.txt