DAILY FANTASY FOOTBALL ROSTER OPTIMIZATION

Maharshi Thakker
John Gearheart
Sakib Salim

OVERVIEW

The challenge of Daily Fantasy Football is to create the highest scoring fantasy football team from a pool of available players given individual player salaries and an overall weekly salary cap (both provided by DFF site)

Previous methods applied to other Daily Fantasy Sports have focused on historical player performance overall and clustering of high-value players as deciding factors for roster formation. Using machine learning we hope to improve upon this methodology for optimizing a DFF roster.

GOAL

Develop a Model for Expected Fantasy Points

Simplest strategy is to optimize budget based on point predictions from DraftKings model (Trust the casino?)

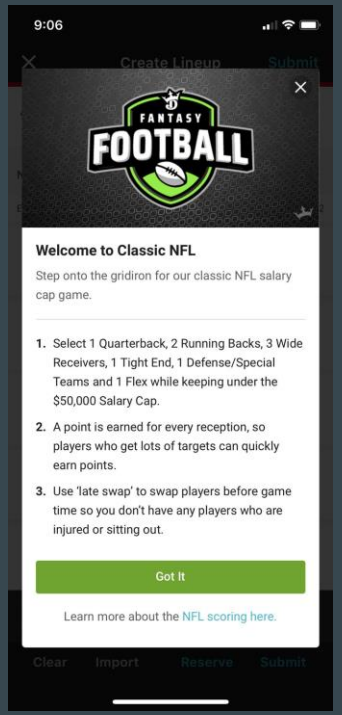
Project intent was to develop a model that minimizes difference between actual points and predicted points and outperform the strategy above

SYSTEM AND METHODS

DraftKings “Classic NFL”

9 Player Lineup
$50,000 Salary Cap

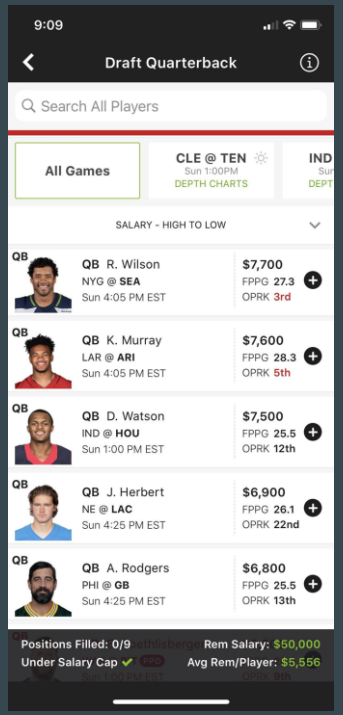
1 Quarterback
2 Running Backs
3 Wide Receivers
1 Tight End
1 Defense
1 Flex Player (RB/WR/TE)



Scoring

Offense	
Passing TD	+4 Pts
25 Passing Yards	+1 Pt (+0.04 Pts/ Yards)
300+ Yard Passing Game	+3 Pts
Interception	-1 Pt
Rushing TD	+6 Pts
10 Rushing Yards	+1 Pt (+0.1 Pts/Yard)
100+ Yard Rushing Game	+3 Pts
Receiving TD	+6 Pts
10 Receiving Yards	+1 Pt (+0.1 Pts/Yard)
100+ Receiving Yard Game	+3 Pts
Reception	+1 Pt
Punt/Kickoff/FG Return for TD	+6 Pts
Fumble Lost	-1 Pt
2 Pt Conversion (Pass, Run, or Catch)	+2 Pts
Offensive Fumble Recovery TD	+6 Pts

Defense	
Sack	+1 Pt
Interception	+2 Pts
Fumble Recovery	+2 Pts
Punt/Kickoff/FG Return for TD	+6 Pts
Interception Return TD	+6 Pts
Fumble Recovery TD	+6 Pts
Blocked Punt or FG Return TD	+6 Pts
Safety	+2 Pts
Blocked Kick	+2 Pts
2 Pt Conversion/Extra Point Return	+2 Pts
0 Points Allowed	+10 Pts
1 – 6 Points Allowed	+7 Pts
7 – 13 Points Allowed	+4 Pts
14 – 20 Points Allowed	+1 Pt
21 – 27 Points Allowed	+0 Pts
28 – 34 Points Allowed	-1 Pt
35+ Points Allowed	-4 Pts

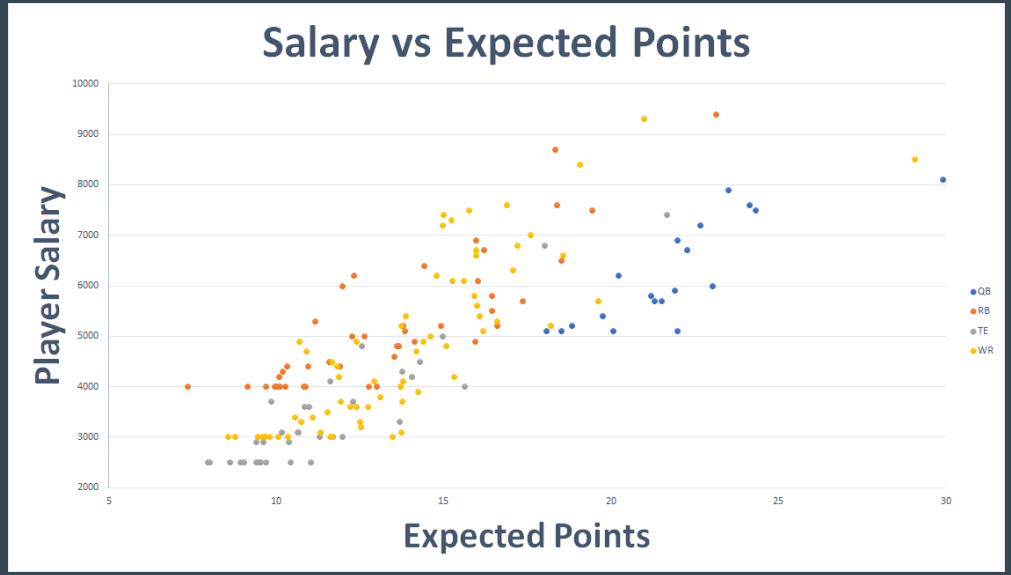


SYSTEM AND METHODS

Data Gathering	• Scrape past years data for each player's stats for each game
Feature Creation	• Normalize point values for each player-game pair • Feature engineer different point categories for each player
Feature Mapping	• From features model expected fantasy points
Optimization	• Select players based on maximum expected points value given cost constraints
Evaluation	• Backtest and compare expected roster to actual season performance

Data Gathering: GOAL

Player	Cost	Expected Fantasy Points
Kyle Juszczyk	4000	8.363361154
Chris Manhertz	2500	8.866679852
Cole Kmet	2900	9.357779652
Kaden Smith	2500	9.422748672
Patrick Laird	4000	9.433700628
Marcedes Lewis	2500	9.627930957
...	...	...



Data Gathering: Available Player Salaries

View Week: 14
Sort first by: Value Factor Second by: Pos Third by: Fantasy Score Go

Player	Pos	year	week	DK SALARY	Fantasy Score	Value Factor	Rank
Seattle, Seahawks	DST	2020	14	$3000	10.4	3.5	1
Tampa Bay, Buccaneers	DST	2020	14	$2900	8.3	2.9	3
Wilson, Russell	QB	2020	14	$7900	22.5	2.9	1
Henry, Derrick	RB	2020	14	$8700	24.6	2.8	1
Gaskin, Myles	RB	2020	14	$5600	15.8	2.8	12
Hunt, Kareem	RB	2020	14	$5000	14.0	2.8	18
Elliott, Ezekiel	RB	2020	14	$6600	18.3	2.8	8

Available Players
Player Costs

DKSalariesScrape.py

SalariesPoints.csv

Available Players +
DK Salary

Data Gathering: Historical Player Performances

STATHEAD FOOTBALL

Query Results Share & more ▾ Glossary

														Passing					
Rk	Tm	Year	Date	Time	LTime		Opp	Week	G#	Day	Result	OT	Cmp	Att	Cmp%	Yds	TD	Int	Sk	Yds	Rate
1	CHI	1943	1943-11-14			@	NYG	9	8	Sun	W 56-7		22	38	57.9	488	7	2	0	0	120.1
2	CLE	1949	1949-10-14			@	LAD	8	7	Fri	W 61-14		17	28	60.7	423	7	1	3	26	129.5
3	DEN	2013	2013-09-05	9:13	7:13		BAL	1	1	Thu	W 49-27		27	42	64.3	445	7	0	3	17	141.1
4	MIN	1969	1969-09-28				BAL	2	2	Sun	W 52-14		36	56	64.3	530	7	2	1	8	120.4
5	NOR	2015	2015-11-01	1:02	12:02		NYG	8	8	Sun	W 52-49		39	50	78.0	505	7	2	0	0	131.7
6	NYG	1962	1962-10-28				WAS	7	7	Sun	W 49-34		27	39	69.2	505	7	0	0	0	151.4
7	HOU	1961	1961-11-19				NYT	11	10	Sun	W 49-13		21	38	55.3	407	7	1	1	20	123.6
8	HOU	1962	1962-10-14				NYT	6	5	Sun	W 56-17		16	28	57.1	252	7	1	1	7	112.9
9	PHI	1954	1954-10-17			@	WAS	4	4	Sun	W 49-21		23	33	69.7	260	7	1	3	29	123.6
10	PHI	2013	2013-11-03	4:05	1:05	@	OAK	9	9	Sun	W 49-20		24	31	77.4	418	7	0	1	1	158.3
11	SDG	1981	1981-11-22	4:00	1:00	@	OAK	12	12	Sun	W 55-21		30	47	63.8	317	7	1	1	7	114.7
12	BUF	1991	1991-09-08	1:00	1:00		PIT	2	2	Sun	W 52-34		31	43	72.1	343	6	2	2	20	117.5
13	CHI	1949	1949-12-11				CRD	12	12	Sun	W 52-21		24	41	58.5	468	6	3	0	0	107.5
14	CHI	2018	2018-09-30	1:00	12:00		TAM	4	4	Sun	W 48-10		19	26	73.1	344	6	0	1	10	154.6

Fantasy Performances
Over Season by Game

PlayerScrape.py

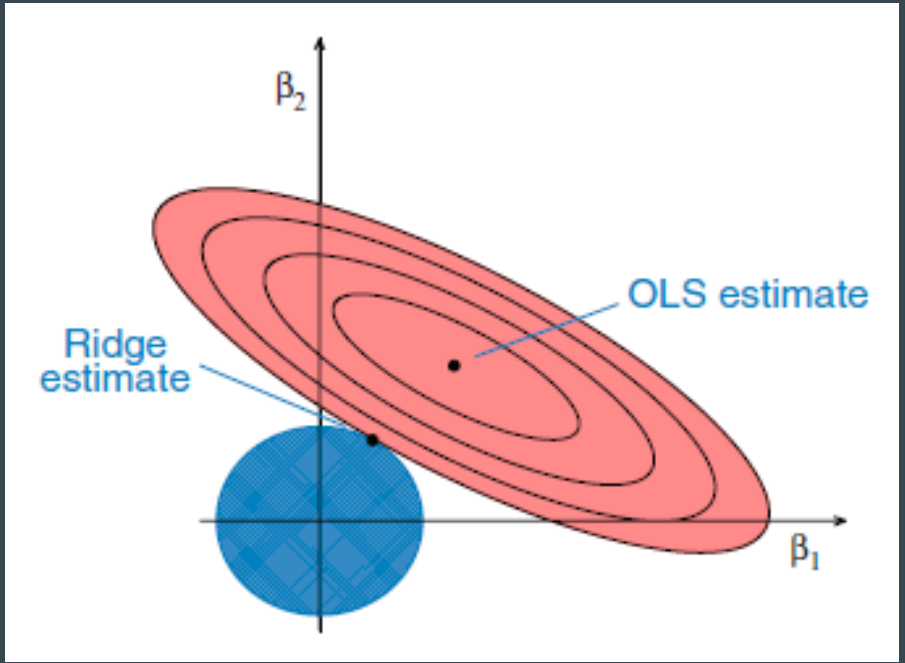
Season<X>ThroughWeek<Y>.csv

All Players + All Fantasy Point
Relevant Stats by Game

Model Selection

Ridge Regression

- We hypothesize there is recency bias to a players performance
- Prevent arbitrary/overweighting to previous game performance



Data Processing

Season<X>ThroughWeek<Y>.csv

All Players + All Stats by Game

Stat1.csv
All Players + Last 5 Performances for a Single Stat

Player	Team	Game N	Game N-1	Game N-2	Game N-3	Game N-4	Position
A.J. Brown	TEN	87	98	62	21	101	WR
A.J. Brown	TEN	98	62	21	101	24	WR
A.J. Brown	TEN	62	21	101	24	153	WR
A.J. Brown	TEN	21	101	24	153	58	WR
A.J. Brown	TEN	101	24	153	58	82	WR
A.J. Brown	TEN	24	153	58	82	39	WR
A.J. Green	CIN	0	0	41	0	19	WR
A.J. Green	CIN	0	41	0	19	82	WR
A.J. Green	CIN	41	0	19	82	96	WR
A.J. Green	CIN	0	19	82	96	0	WR

Feature Creation

Player	Team	Game N	Game N-1	Game N-2	Game N-3	Game N-4	Position
A.J. Brown	TEN	87	98	62	21	101	WR
A.J. Brown	TEN	98	62	21	101	24	WR
A.J. Brown	TEN	62	21	101	24	153	WR
A.J. Brown	TEN	21	101	24	153	58	WR
A.J. Brown	TEN	101	24	153	58	82	WR
A.J. Brown	TEN	24	153	58	82	39	WR
A.J. Green	CIN	0	0	41	0	19	WR
A.J. Green	CIN	0	41	0	19	82	WR
A.J. Green	CIN	41	0	19	82	96	WR
A.J. Green	CIN	0	19	82	96	0	WR

First Attempt

- Predict 5th game's Fantasy Points based on past 4 rolling games for each player

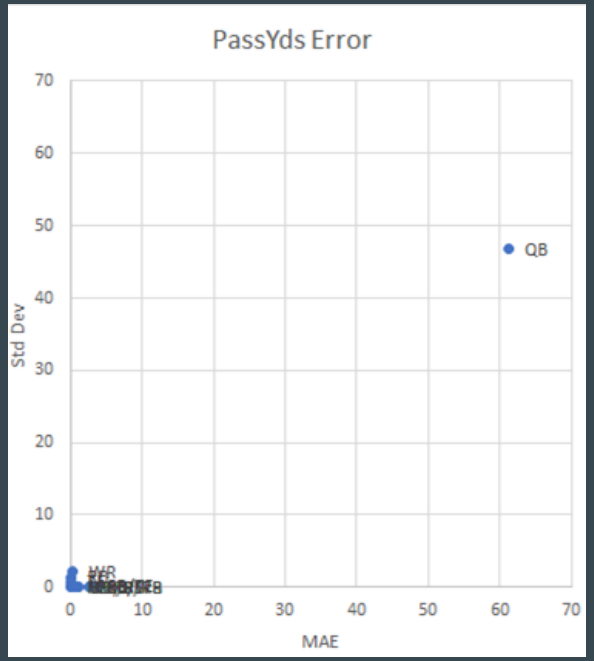
Second Attempt

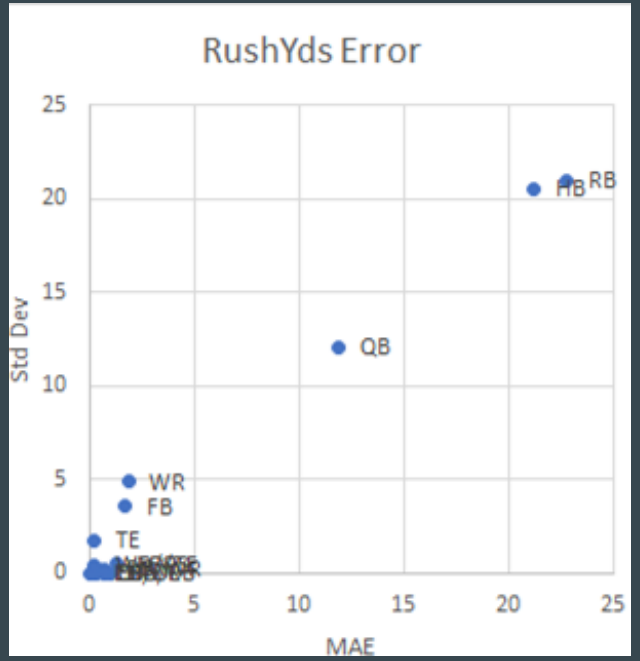
- Predict 5th game's individual stats (PassYds, PassTD, etc.) based on past 4 rolling games for each player

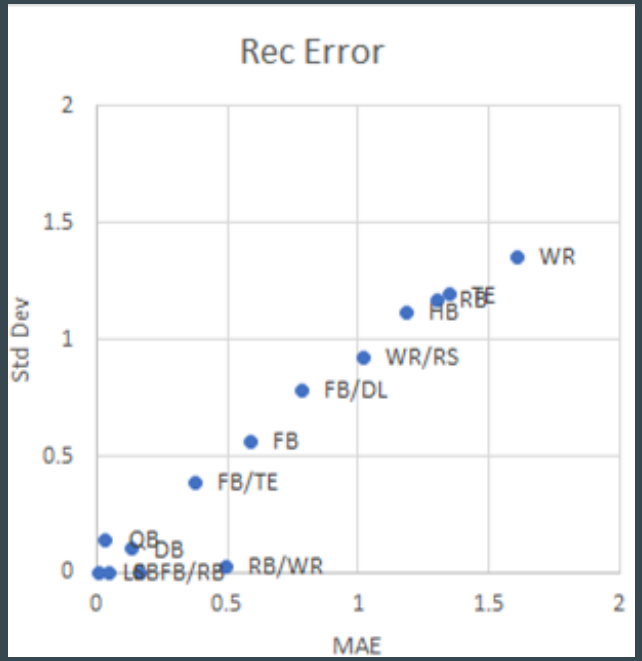
Third Attempt

- Predict 5th game's individual stats based on past 4 rolling games for each player and position

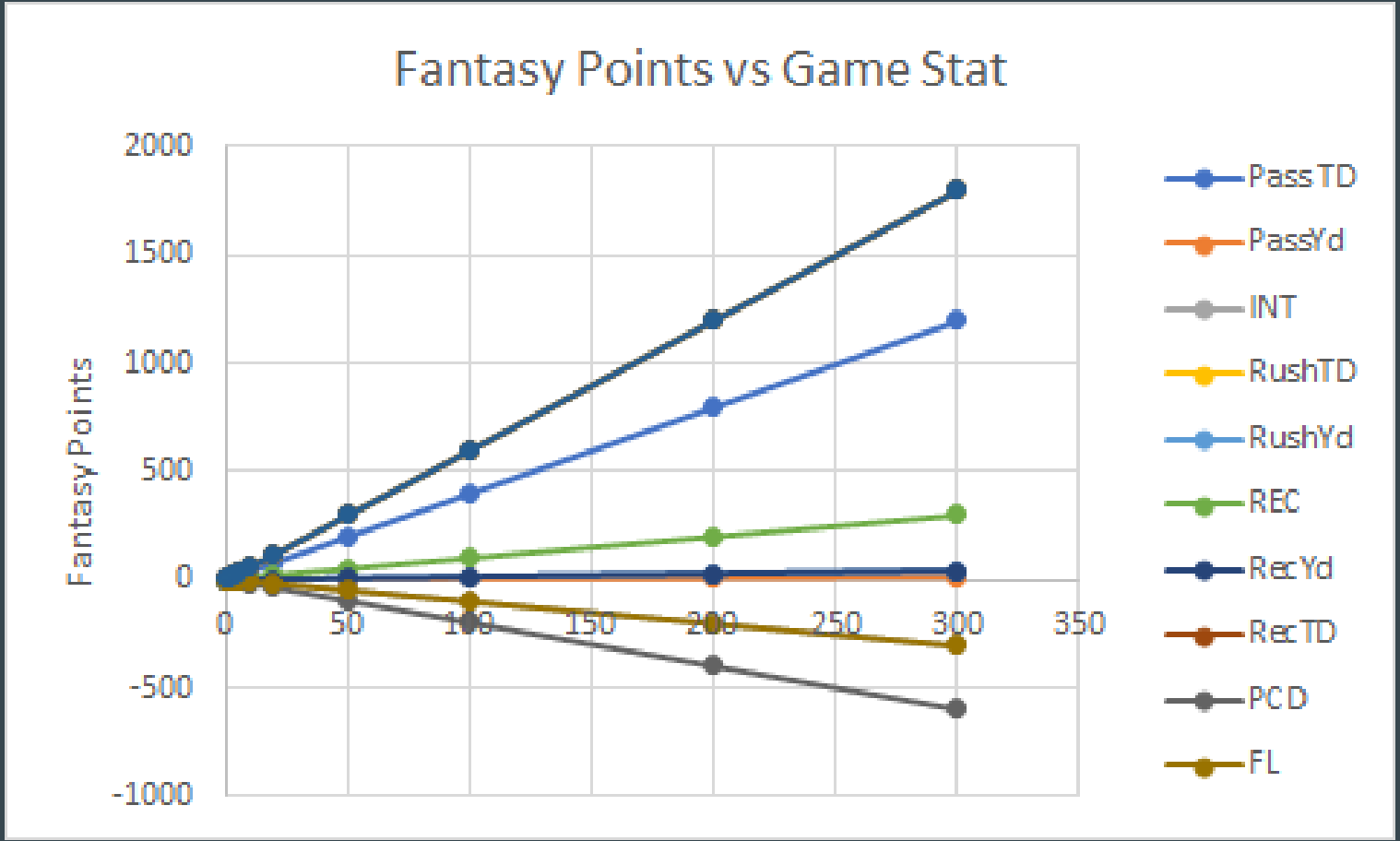
Model Performance







Model Sensitivity



Data Processing

Available Players + Costs.csv

Predicted Points.csv

Player	Cost	Expected Fantasy Points
Kyle Juszczyk	4000	8.363361154
Chris Manhertz	2500	8.866679852
Cole Kmet	2900	9.357779652
Kaden Smith	2500	9.422748672
Patrick Laird	4000	9.433700628
Marcedes Lewis	2500	9.627930957
...	...	...

Roster Optimization Input Data

Optimization - Integer Program

- For a set of players P, we can define disjoint subsets: Q, W, R, T, D to represent the different positions. We also define variables for points and salaries.
- Constraints
    - Rostered # of Players by Position
    - Total # of Players
    - Total Roster Salary
- We formulate the IP:

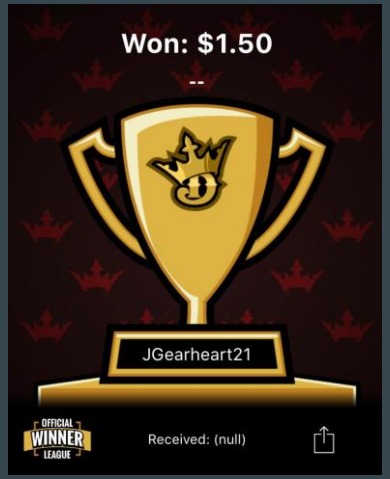
$$
\begin{aligned}
\text{maximize} \quad & \sum f_i x_i \\
\text{subject to} \quad & \sum s_i x_i \leq 50000 \\
& \sum_{p_i \in Q} x_i = 1 \\
& 3 \leq \sum_{p_i \in W} x_i \leq 4 \\
& 2 \leq \sum_{p_i \in R} x_i \leq 3 \\
& 1 \leq \sum_{p_i \in T} x_i \leq 2 \\
& \sum_{p_i \in D} x_i = 1 \\
& x_i \in \{0, 1\}
\end{aligned}
$$

Optimal Lineup and BackTesting

- Use model predictions and DK salaries to solve IP
- For any given week we can compare lineups based on four methodologies:
    - Model Predictions
    - DK Projections
    - Aggregated “Expert” Projections (FantasyPros)
    - True Fantasy Points
- Billions of possible lineups: perfect lineup is not feasible

Results

- Week 14 Results - Out of Potential 260.14 points:
  - DraftKings: 141.84 (ExpectedPts - ActualPts = 9.15)
  - Our Model: 141.52 (ExpectedPts - ActualPts = 28.73)
  - FantasyPros: 133.54 (ExpectedPts - ActualPts = 28.03)



+50% return
-DK and Our Model Paid out
-No winnings for FP team

- On par with DK for Actual Points and similar model accuracy to a commercial Fantasy projection website
- From the best potential roster - Our model picked out two of the optimal players

Results Across 2020

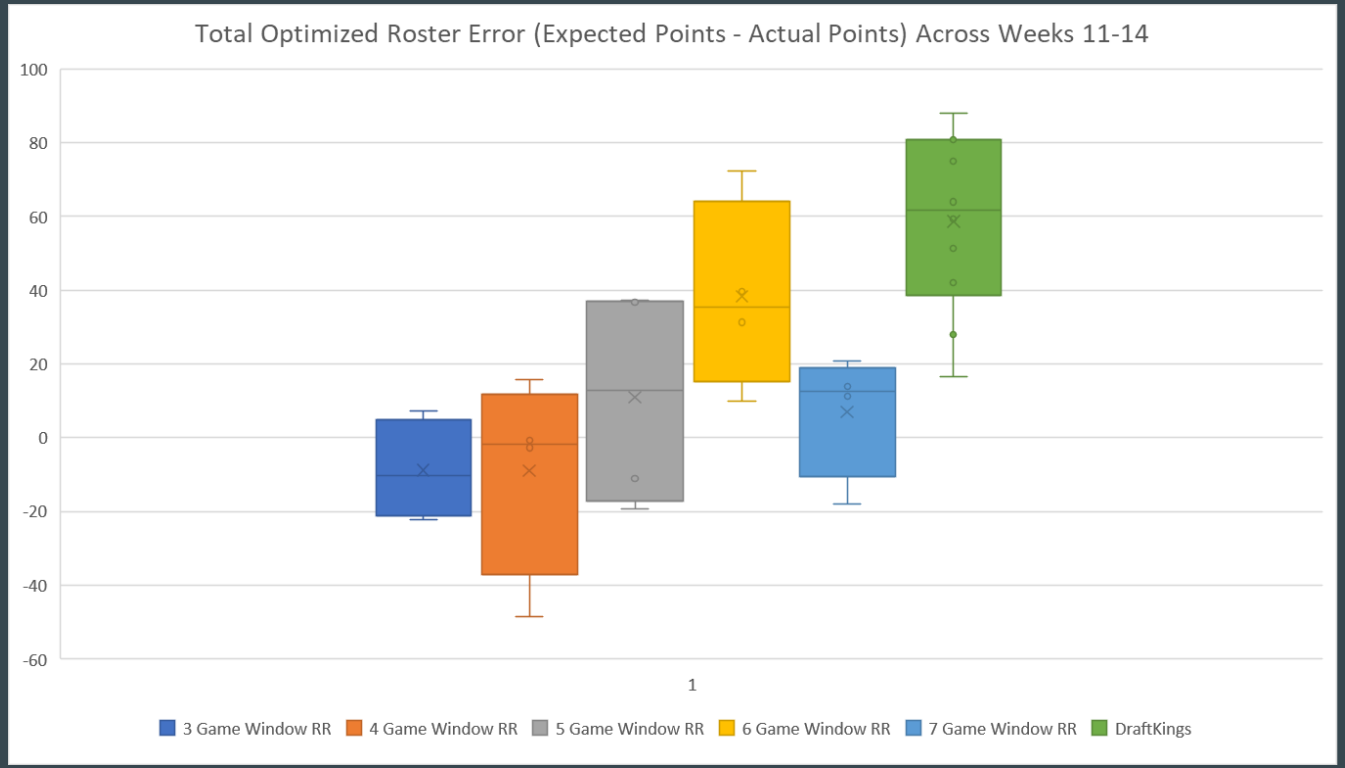
		Ridge-Regression Roster Actual Score									
		Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
Ridge-Regression Window Size	7							138.48	160.5	133.42	144.44
	6					99.1	101.58	136.68	157.24	100.72	127.34
	5			86.32	117.74	97.1	147.1	139.68	140.12	93.12	141.22
	4		109.94	127.2	94.82	84.9	96	128.78	173.72	136.12	111.66
	3	106.32	103.6115	104.22	103.12	86.8	149.7	145.56	112	126.72	139.74

	Optimized Roster Actual Score									
	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
DraftKings Model	115.33	102.74	124.56	91.72	150.18	85.32	85.88	107.34	78.66	138.66

		Ridge-Regression Actual Score > DraftKings ? (1 = yes, 0 = no)									
		5	6	7	8	9	10	11	12	13	14
Ridge-Regression Window Size	7							1	1	1	1
	6					0	1	1	1	1	0
	5			0	1	0	1	1	1	1	1
	4		1	1	1	0	1	1	1	1	0
	3	0	1	0	1	0	1	1	1	1	1

Ridge-Regression Window Length	Win % Over DraftKings
7 Games	100%
6 Games	67%
5 Games	75%
4 Games	78%
3 Games	70%

Results Across 2020



Future Improvements

Feature Selection

- Additional features: long game history, home vs away, opposing team strength vs position, mutual-success player stacking (QB-WR)

Ensemble Method

- Explore different models for outlier positions

Model Selection

- Gradient Boosting
- Neural Network

Contest Selection

THANK YOU*

*This Content is for informational purposes only, you should not construe any such information or other material as legal, investment, financial, or other advice. Individual Results may vary.

The Team

Maharshi Thakker

Sakib Salim

John Gearheart