Optimizing Daily Fantasy Football

MGSC434 - Professor Rim Hariss

Written By

Maëlys Boudier
Jonathan Steinberg

McGill University
Canada

April 11, 2022

1 Introduction

Do you want to become a sports fan but can't decide on a team to cheer for? Were you never athletic enough to compete in professional sports? Or, do you like sports betting but are tired of your ten-leg parlays not hitting? Do not fret; the answer to all these questions is Fantasy Sports. Fantasy sports, a 7.2 billion dollar industry, is a virtual game where users build an imaginary lineups composed of real players [Willingham, 2020]. Fantasy points are allocated based on the real-life performance of players, and typically the lineup with the highest sum of fantasy points at the end of the season wins. Fantasy football is by far the most popular fantasy sport with around 35 million users in the United States, accounting for a whopping 78% of all fantasy sports users [Willingham, 2020]. A very popular style of fantasy football is called daily fantasy football, in which competitions occur on a much shorter horizon; users draft a new lineuo every week instead of keeping the same one for the entirety of the season. In this fast-paced format, users must stay informed on the latest trends and news in order to build high performing lineups. In fact, On DraftKings, a popular platform for daily fantasy football, it is estimated that highly experienced users reap 76% of all distributed winnings, while the remaining 24% go to all other users [DraftsKings, 2022]. DraftKings hosts two types of competitions: GPP Games, in which a large portion of winnings is reserved for the best lineups, and Cash Games, where winnings are distributed evenly amongst the top 50% of lineups [Gresham, 2020]. In this project, we will explore whether an ample amount of football knowledge is required to be successful in Cash Game formats, or whether you can rely on data analytics (and a little amateur football knowledge) to produce above average daily fantasy football lineups.

2 Methodology

To create above average lineups capable of winning in a fantasy football Cash Games, we first built a regression model for each position using 2019 and 2020 data to predict how many fantasy points each player will score for games during the 2021 season. Then, we built an optimization model to determine the best lineup for each week during the 2021 season. Positions are as follows: Quarterback (QB), running back (RB), wide receiver (WR), tight end (TE), and defence (DST).

2.1 Data Overview

Our quest to become fantasy football phenoms started by collecting player statistic data from FantasyData.com and injury data from SportsData.io. This provided us with fantasy points, salaries, injuries, and typical game-by-game statistics for every player in the NFL during the 2019, 2020, and 2021 seasons. However, we felt that the raw data was insufficient to build a robust model to predict fantasy points, and that feature engineering was required.

We took some inspiration from Michael Salfino, a fantasy football writer for The Athletic, who produces a market share report each week specifying players' usage during their previous games [Salfino, 2019]. The idea here is that the higher one's usage, the more opportunities they have to score fantasy points, thus they should be incorporated into lineups. We attempted to capture this relationship by creating the following variables:

- Pass Share: Passing attempts of player ÷ Team's total offensive plays (created for QBs)
- Rush Share: Rushing attempts of player ÷ Team's total running plays (created for RBs)
- Target Share: Intended targets of player ÷ Team's total passing plays (created for WRs)

Obviously, we cannot use a player's statistics to predict how many fantasy points they will score in the same game. Therefore, we computed three types of lagged averages for fantasy points and the share variables above; a rolling average of all previous games, an average of the last five games, and an average of the last three games. We opted to create lagged averages as supposed to taking the lags of the statistics themselves to account for player absences due to injuries or bye weeks.

The variables that we have created thus far are all player-centric, but a player's fantasy production is also highly dependent on the opponent they face. For example, a RB playing against a defense that specializes at stopping the run may gain less points than he usually does. Luckily, our raw dataset contained an Upcoming Opponent Position Rank (OPR) variable for each player. OPR ranks each opponent based on their historical performance against a player's position, where lower numbers correspond to tougher opponents. Incorporating this variable into our predictive models will hopefully place a premium on players who are facing easier opponents, just like an experienced fantasy player would do when building their lineups.

Injured teammates can also play a vital role in a player's fantasy production. For example, the second-best WR may thrive if the best WR on their team is injured. On the other hand, if a WR's QB is injured, his fantasy points may take a hit. For each player in our dataset, we created four injury variables: Two of which were binary variables indicating whether the starting QB and TE on a player's team are injured, respectively. The other two variables counted the amount of fantasy-relevant RBs and WRs that were injured on a player's team, respectively. We defined fantasy-relevant players as those who's rolling average of fantasy points are higher than their team's average.

A exhaustive description of each variable and which positions they correspond to can be found in Appendix 1.

2.2 Prediction Model

We took an iterative approach in constructing the predictive models for each position. First, we ran simple linear regressions on all predictors to get a sense of which had the most significant relationship with fantasy points. We also analyzed a correlation matrix of our predictors for each position. Not surprisingly, the three lagged averages for the same statistic were all highly correlated with each other, so we opted to only include one of each type in our models. To select the best type of averages, we ran regression models with every possible combination of fantasy points and the relevant share variable for that position. For example, target share is the relevant share variable for the WRs, so we iterated through models containing the three types of lagged averages for fantasy points and target share. We selected the combination of the two predictors that were most significant and produced the lowest MSE in our validation dataset. Then, we recursively added the remaining predictors to the model and kept them in if they were significant or improved out-of-sample performance. Finally we attempted to use black-box methods such as Random Forest and Gradient Boosting to improve our prediction accuracy, however multiple linear regression was found to be the best model for all positions.

Once we finalized a regression model for each position, we used them to predict player's fantasy points for the 2021 season. Our predictions achieved a MSE of 38.59 across all players, which corresponds to a deviation of +/- 6.21 fantasy points on average. While this may seem subpar, our predictions in comparison to DraftKings' official projections were only 1.86 points off on average during the 2021 season. This suggests that our models are nearing their potential, and that the high MSE of realized fantasy points is due to the unpredictable nature of the game. The MSEs for each model can be found in Appendix 8.

2.3 Optimization Model

Our goal was to maximize the number of fantasy points subject to the rules imposed by DraftKings. The model appears as follows:

Decision Variables: Which players to draft to our lineup

$$X_i = \begin{cases} 1 & \text{If we draft the player to our lineup} \\ 0 & \text{Otherwise} \end{cases}$$

Other Variables:

P_i = Fantasy points of player i
C_i = Salary of player i (Cost)
$QB_i = 1$ if player i is a QB, 0 otherwise
$WR_i = 1$ if player i is a WR, 0 otherwise
$TE_i = 1$ if player i is a TE, 0 otherwise
$DST_i = 1$ if team i is DST, 0 otherwise

Objective Function: Maximize fantasy points

$$maxZ = \sum_{i=1} P_i X_i$$

Base Constraints: Rules imposed by DraftKings

1. **Monetary Constraints:** Budget of \$50,000: $\sum_{i=1} C_i X_i \leq 50000$
2. **Positional Constraints:**

   Note: Lineup can have one flexible player (FLEX) who can be an extra RB, WR, or TE
   1 QB: $\sum_{i=1} QB_i X_i = 1$
   2 RB (or 3 if FLEX): $2 \leq \sum_{i=1} QB_i X_i \leq 3$
   3 WR (or 4 if FLEX): $3 \leq \sum_{i=1} QB_i X_i \leq 4$
   1 TE (or 2 if FLEX): $1 \leq \sum_{i=1} QB_i X_i \leq 2$
   1 DST: $\sum_{i=1} DST_i X_i = 1$
   9 players per lineup: $\sum_{i=1} X_i = 9$

In keeping with the spirit of our project, we wanted to introduce additional constraints to improve the performance of our lineups in a data-driven fashion. We used realized fantasy points from the 2019 and 2020 seasons to build 32 perfect-hindsight lineups. Essentially, these are the best possible lineups that could have been constructed for each week of the previous two seasons. Out of 32 lineups we found that only three contained players that were playing against the selected DST; this makes sense as they would otherwise be cannibalizing each other's points. In addition, only two lineups had a QB and a RB who are teammates in real life. Intuitively, this makes sense; teams are either pass or run-oriented depending on their style of offense and situation of the game. There is often a trade-off to be made and it is very rare that a team's QB and RB will score a lot of fantasy points in the same game. Furthermore, 20 of the perfect lineups had a WR in their FLEX position. This is because DraftKings operates a Point-Per-Reception (PPR) format, in which players gain an additional fantasy point every time they catch a pass. Surprisingly, only ten lineups had a QB and a WR who are on the same team in real life. This tactic, known as Stacking, is often used by fantasy football users because QBs and WRs can both gain fantasy points on the same play. Therefore, if the QB scores a lot of fantasy points, it is likely that the WR will too, and vice-versa. However, our perfect-hindsight lineups fail to convince us that this is a successful strategy, so we opted to not include this constraint in our optimization model (Appendix 11). The additional constraints are as follows:

New Variables:
$Team_j, i = 1$ if player i belongs to team j, 0 otherwise
$Opponent_j, i = 1$ if player i plays against team j, 0 otherwise

Additional Constraints: Derived from perfect-hindsight lineups

1. QB and RB must be on different teams: $2 * \sum iX_i * QB_i * Team_j, i \geq \sum iX_i * RB_i * Team_j, i \forall j$

2. Players cannot play against DST
   QB: $1 \geq \sum iX_i * DST_i * Team_j, i + X_i * QB_i * Opponent_j, i \ \forall j$
   TE: $1 \geq \sum iX_i * DST_i * Team_j, i + X_i * QB_i * Opponent_j, i \ \forall j$
   RB: $2 * (1 - \sum iX_i * DST_i * Team_j, i) \geq \sum iX_i * RB_i * Opponent_j, i \ \forall j$
   WR: $4 * (1 - \sum iX_i * DST_i * Team_j, i) \geq \sum iX_i * RB_i * Opponent_j, i \ \forall j$

3. FLEX must be a WR: $\sum_{i=1} WR_i X_i = 4$

Once we incorporated the additional constraints, we optimized our model to construct the best possible lineups for weeks 6-17 during the 2021 season using our fantasy point predictions. We excluded the first five weeks to let our averages of the last five games be calculated and excluded the last week of the season as many teams rest their starters in anticipation for the playoffs.

3 Results and Conclusion

Our twelve optimization models achieved an average of 115.8 fantasy points. Our worst score of 94.28 points came in week 7 while our best score of 164.58 points came in week 13. Interestingly, our additional constraints had a negligible impact on our performance; these lineups achieved an average of 116.3 fantasy points and never scored less points than our original lineups. A possible explanation for this is that our fantasy point predictions accounted for some of the constraints in our model. For example, the average predicted fantasy points for WRs were higher than that of RBs or TEs, therefore our optimization model would naturally include a WR in the FLEX position (Appendix 10).

In comparison to lineups that won DraftKings' Fantasy Football Millionaire Tournament, our lineups earned 112.8 less points on average [Eric Lindquist, 2021] (Appendix 9). The closest we came to a winning lineup was in week 7, where we were only 77.5 points away from racking in a wad of cash. While the average spread is quite large, benchmarking ourselves against these lineups is unfair as users often need to rely more on luck rather than skill in order to achieve the best lineup in a competition.

In order to truly assess whether we are able to outperform a knowledgeable daily fantasy football user, we bench-marked ourselves against Chris Pennell, an avid DraftKings user who publishes his lineups each week on his YouTube channel to his audience of 9,200 subscribers [Pennell, 2021]. Unfortunately, our lineups achieved 14.6 less points on average compared to Chris' lineups, only outright beating him in three out of twelve weeks (Appendix 9). Perhaps our lineups are good enough to win most weeks in Cash Games, but they fail to beat knowledgeable players like Chris, which makes sense in such a highly nuanced game like fantasy football. Users like Chris who follow the NFL closely are able to rely on their intuition in drafting successfully their lineups, giving them a competitive advantage against our objective approach. However, our model can still serve as a gut-check to even the most experienced players, and can certainly guide beginners who lack experience. After all, its your own money that you will be risking; you might sleep better at night if you go with your gut.

References

[DraftsKings, 2022] DraftsKings (2022). Average Results. *DraftsKings Inc.* https://www.draftkings.com/average-results.

[Eric Lindquist, 2021] Eric Lindquist, G. E. (2021). DraftsKings Picks. https://www.youtube.com/channel/UCBUUOHOLmDuWii7zBV7SKrQ.

[Gresham, 2020] Gresham, C. (2020). Beginner NFL DFS: GPP vs. Cash. *DKNation.* https://dknation.draftkings.com/2020/6/10/21284815/daily-fantasy-sports-nfl-dfs-beginner-definitions-glossary-gpp-cash-strategy.

[Pennell, 2021] Pennell, C. (2021). DraftsKings Picks & Strategy. https://www.youtube.com/user/cpennell16.

[Salfino, 2019] Salfino, M. (2019). Market share report: A banner week for the zeroRB crowd, Rashod Bateman earns instant respect and more. *The Athletic.* https://theathletic.com.

[Willingham, 2020] Willingham, A. (2020). Fantasy football is a billion-dollar pastime. Covid-19 is wreaking havoc with it. *CNN.* https://cnn.com.

Appendix

Variable List

Variable	Description	Applicable Positions
FantasyPointsMean_Rolling	*Average fantasy points of all previous games of player X*	QB, RB, WR, TE
FantasyPointsMean_Last5	*Average fantasy points of last five games of player X*	QB, RB, WR, TE
FantasyPointsMean_Last3	*Average fantasy points of last three games of player X*	QB, RB, WR, TE
PassShareMean_Rolling	*Average pass share of all previous games of player X*	QB
PassShareMean_Last5	*Average pass share of last five games of player X*	QB
PassShareMean_Last3	*Average pass share of last three games of player X*	QB
RushShareMean_Rolling	*Average rush share of all previous games of player X*	RB, WR, TE
RushShareMean_Last5	*Average rush share of last five games of player X*	RB, WR, TE
RushShareMean_Last3	*Average rush share of last three games of player X*	RB, WR, TE
TargetShareMean_Rolling	*Average target share of all previous games of player X*	RB, WR, TE
TargetShareMean_Last5	*Average target share of last five games of player X*	RB, WR, TE
TargetShareMean_Last3	*Average target share of last three games of player X*	RB, WR, TE
FantasyPointsForMean_Rolling	*Average fantasy points of all previous games of defence X*	DST
FantasyPointsForMean_Last5	*Average fantasy points of last five games of defence X*	DST
FantasyPointsForMean_Last3	*Average fantasy points of last three games of defence X*	DST
FantasyPointsAgainstMean_Rolling	*Average fantasy points of all previous games that defences score when playing a opponent X*	DST
FantasyPointsAgainstMean_Last5	*Average fantasy points of last 5 gamesthat defences score when playing a opponent X*	DST
FantasyPointsAgainstMean_Last3	*Average fantasy points of last 3 games that defences score when playing a opponent X*	DST
OPR (Opponent Position Rank)	*Rank of opponent based on their historical performance against player X's position*	QB, RB, WR, TE
QBInjured	*Whether the starting quarterback on player X's team is injured*	QB, RB, WR, TE
RBInjured	*Amount of fantasy-relevant running backs that are injured on player X's team*	QB, RB, WR, TE
WRInjured	*Amount of fantasy-relevant wide receivers that are injured on player X's team*	QB, RB, WR, TE
TEInjured	*Whether the starting tight end on player X's team is injured*	QB, RB, WR, TE
QBInjured	*Whether the starting quarterback on defence X opponent's team is injured*	DST
RBInjured	*Amount of fantasy-relevant running backs that are injured on defence X opponent's team*	DST
WRInjured	*Amount of fantasy-relevant wide receivers that are injured on defence X opponent's team*	DST
TEInjured	*Whether the starting tight end on defence X opponent's team is injured*	DST

Figure 1: All Variables

WR Regression Model

	Dependent variable:
	FantasyPoints
FantasyPointsMean_Rolling	0.548***
	(0.055)
TargetShareMean_Last3	22.315***
	(3.590)
OPR	0.042**
	(0.017)
QBInjured	-1.692***
	(0.630)
WRInjured	1.501***
	(0.395)
Constant	-0.250
	(0.408)
Observations	1,868
R^2	0.362
Adjusted R^2	0.361
Residual Std. Error	6.967 (df = 1862)
F Statistic	211.585*** (df = 5; 1862)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2: WR Regression

RB Regression Model

	Dependent variable:
	FantasyPoints
FantasyPointsMean_Rolling	0.635***
	(0.047)
RushShareMean_Last3	6.883***
	(1.132)
OPR	0.068***
	(0.018)
RBInjured	0.926**
	(0.398)
WRInjured	0.740**
	(0.376)
Constant	-0.768*
	(0.394)
Observations	1,383
R^2	0.483
Adjusted R^2	0.481
Residual Std. Error	6.232 (df = 1377)
F Statistic	257.151*** (df = 5; 1377)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 3: RB Regression

TE Regression Model

	Dependent variable:
	FantasyPoints
FantasyPointsMean_Last3	0.302^{***}
	(0.052)
TargetShareMean_Rolling	38.767^{***}
	(3.770)
QBInjured	-1.093^{**}
	(0.530)
Constant	0.608^{***}
	(0.216)
Observations	1,097
R^2	0.424
Adjusted R^2	0.422
Residual Std. Error	4.654 (df = 1093)
F Statistic	267.644^{***} (df = 3; 1093)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 4: TE Regression

QB Regression Model

	Dependent variable:
	FantasyPoints
FantasyPointsMean_Last5	0.502^{***}
	(0.091)
PassShareMean_Last3	7.722^{**}
	(3.669)
OPR	0.083^{*}
	(0.046)
Constant	3.450^{*}
	(1.895)
Observations	411
R^2	0.170
Adjusted R^2	0.164
Residual Std. Error	8.733 (df = 407)
F Statistic	27.876^{***} (df = 3; 407)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 5: QB Regression

DST Regression Model

	Dependent variable:
	FantasyPoints
FantasyPointsFor_MeanLast3	0.205^{***}
	(0.076)
FantasyPointsAgainst_MeanRolling	0.441^{***}
	(0.106)
Constant	2.856^{***}
	(0.919)
Observations	441
R^2	0.053
Adjusted R^2	0.049
Residual Std. Error	6.098 (df = 438)
F Statistic	12.314^{***} (df = 2; 438)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 6: DST Regression

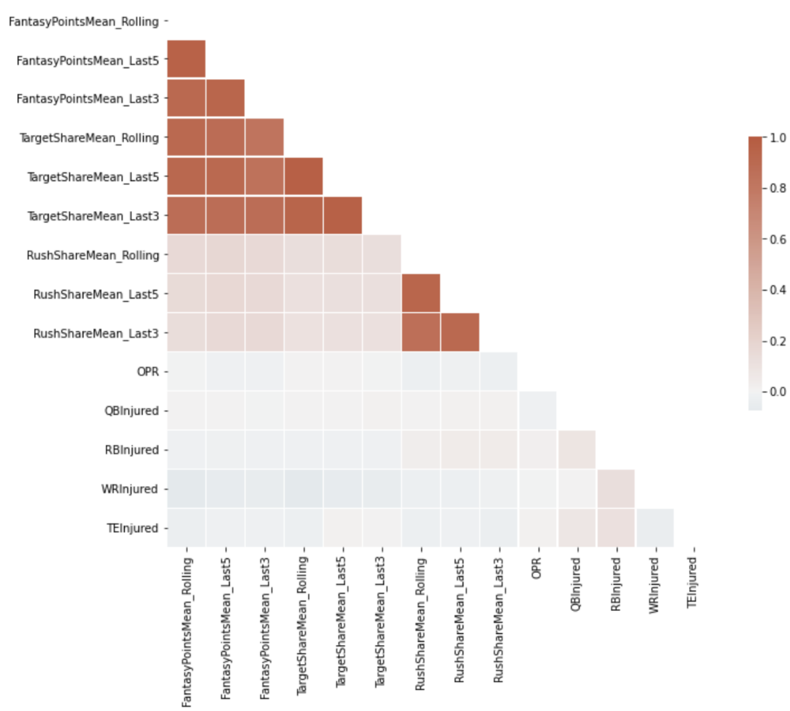

Figure 7: Correlation Matrix

Position	Fantasy Points		DraftKings Projection	
	MSE	*Deviation*	*MSE*	*Deviation*
QB	67.38	8.21	11.60	3.41
RB	41.63	6.45	4.55	2.13
WR	40.91	6.40	2.52	1.59
TE	21.66	4.65	1.69	1.30
DST	36.10	6.01	1.25	1.12
Average	**38.81**	**6.23**	**3.30**	**1.82**

Figure 8: Regression Results

Week	Our Predictions (Constraints)	Our Predictions (Rules)	Chris Pennell	Winning Lineups
6	112.1	114.3	133.1	NaN
7	94.3	94.3	113.2	239.1
8	103.1	103.1	144.6	226.2
9	124.3	124.5	118.9	205.3
10	106.1	106.1	130.9	213.1
11	98.7	98.7	93.8	228.2
12	124.0	124.0	100.5	225.3
13	164.6	164.6	168.8	242.1
14	131.1	131.1	142.1	222.8
15	94.6	94.6	118.8	222.4
16	117.0	121.0	143.9	268.5
17	119.7	119.7	155.8	214.5
Average	**115.8**	**116.3**	**130.4**	**228.0**

Figure 9: Optimization Results

Average Predicted Fantasy Points	
Position	*Points*
QB	18.00
RB	7.58
WR	7.32
TE	4.70
DST	7.02

Figure 10: Points/Position

Amount of Perfect-Hindsight Lineups that Contain…	
QB & RB on same team	2
QB & WR on same team	12
QB & TE on same team	3
RB & WR on same team	3
RB & TE on same team	1
WR & TE on same team	3
DST is QB opponent	0
DST is RB opponent	1
DST WR opponent	0
DST is TE opponent	2
FLEX is RB	10
FLEX is WR	20
FLEX is TE	2

Figure 11: Perfect Hindsight Trends (2019-2020)