Who do I start?

How can past performance help predict future performance in NFL fantasy football

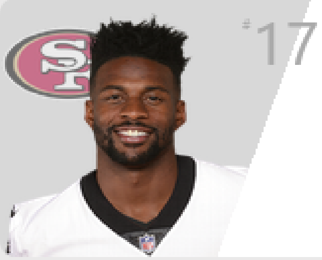
Rory Breslin

- Springboard Capstone Project (May 2020 Cohort)
- Thanks to Springboard mentor Max Sopp

Week 7 of your season and you are already in desperate need for a win…which 3 WRs do you start?

#11 **Robby Anderson**
New York Jets
ELIG WR
MANAGER Tmp
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	Dal	5	125	25.0	1	8	0	0	0	0	20.5
5	@Phi	1	16	16.0	0	3	0	0	0	0	1.6
4	BYE	--	--	--	--	--	--	--	--	--	--
3	@NE	3	11	3.7	0	5	0	0	0	0	1.1
2	Cle	4	81	20.3	0	6	0	0	0	0	8.1
1	Buf	3	23	7.7	0	7	0	0	0	0	2.3



#17 **Emmanuel Sanders**
San Francisco 49ers
ELIG WR
MANAGER BRES
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	Ten	1	0	0.0	0	3	0	0	0	0	0.0
5	@LAC	1	9	9.0	0	1	0	0	0	0	0.9
4	Jax	5	104	20.8	0	9	0	0	0	0	12.4
3	@GB	2	10	5.0	0	5	0	0	0	0	1.0
2	Chi	11	98	8.9	1	13	0	0	0	0	17.8
1	@Oak	--	--	--	--	--	--	--	--	--	--

#19 **Adam Thielen**
Minnesota Vikings
ELIG WR
MANAGER BRES
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	Phi	6	57	9.5	1	8	1	0	0	0	11.7
5	@NYG	7	130	18.6	2	8	0	0	0	0	27.0
4	@Chi	2	6	3.0	0	6	0	0	0	0	0.6
3	Oak	3	55	18.3	1	5	1	1	1	0	17.6
2	@GB	5	75	15.0	0	8	0	0	0	0	7.5
1	Atl	3	43	14.3	1	3	0	0	0	0	10.3

#83 **Marquez Valdes-Scantling**
Green Bay Packers
ELIG WR
MANAGER Free Agent
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	Det	2	48	24.0	0	2	1	9	0	0	5.7
5	@Dal	1	18	18.0	0	4	0	0	0	0	1.8
4	Phi	3	47	15.7	0	7	0	0	0	0	4.7
3	Den	6	99	16.5	1	10	0	0	0	0	15.9
2	Min	3	19	6.3	0	5	0	0	0	0	1.9
1	@Chi	4	52	13.0	0	6	1	0	0	0	5.2

#83 **Tyler Boyd**
Cincinnati Bengals
ELIG WR
MANAGER BRES
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	@Bal	3	10	3.3	0	7	0	0	0	0	1.0
5	Ari	10	123	12.3	1	14	0	0	0	0	20.3
4	@Pit	3	33	11.0	0	6	0	0	0	0	3.3
3	@Buf	6	68	11.3	0	11	0	0	0	0	6.8
2	SF	10	122	12.2	0	10	0	0	0	0	14.2
1	@Sea	8	60	7.5	0	11	1	3	0	0	6.3

#84 **Corey Davis**
Tennessee Titans
ELIG WR
MANAGER Free Agent
STATUS Healthy

WEEK	OPP	REC	YDS	AVG	TD	TAR	CAR	YDS	TD	MISCTD	FPTS
6	@Den	3	36	12.0	0	5	0	0	0	0	3.6
5	Buf	2	28	14.0	0	4	0	0	0	0	2.8
4	@Atl	5	91	18.2	1	6	0	0	0	0	15.1
3	@Jax	3	44	14.7	0	4	0	0	0	0	4.4
2	Ind	3	38	12.7	0	5	0	0	0	0	3.8
1	@Cle	0	0	0.0	0	3	0	0	0	0	0.0

Every week fantasy owners have to decide who to start on their team to give them the best chance of winning

WINNER **136.2** 122.5 LOSER

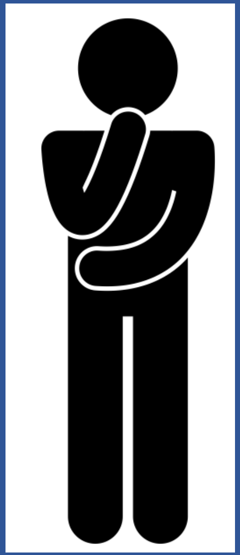
Overview | Full Box Score | Scoring Breakdown

STARTERS		NFL WEEK 10			TOTAL
SLOT	PLAYER	OPP	STATUS	PROJ	FPTS
QB	Kyler Murray Ari QB	@TB	L 27-30	20.8	33.6
RB	Josh Jacobs Oak RB	LAC	W 26-24	16.4	16.1
RB	J.D. McKissic Det RB	@Chi	L 13-20	4.5	5.5
WR	Allen Robinson II Chi WR	Det	W 20-13	8.4	8.6
WR	Curtis Samuel Car WR	@GB	L 16-24	8.6	9.5
TE	Hunter Henry LAC TE	@Oak	L 24-26	11.3	9.0
FLEX	Jamison Crowder NYJ WR	NYG	W 34-27	7.5	14.1
FLEX	Greg Olsen Sea TE	@SF	W 27-24	5.7	9.8
D/ST	49ers D/ST SF D/ST	Sea	L 24-27	7.1	23.0
K	Justin Tucker Bal K	@Cin	W 49-13	8.7	7.0
		TOTALS		98.9	136.2

STARTERS		NFL WEEK 10			TOTAL
SLOT	PLAYER	OPP	STATUS	PROJ	FPTS
QB	Ryan Tannehill Ten QB	KC	W 35-32	17.7	20.9
RB	Dalvin Cook Min RB	@Dal	W 28-24	17.3	24.3
RB	Mark Ingram II Bal RB	@Cin	W 49-13	14.8	9.4
WR	John Brown Buf WR	@Cle	L 16-19	8.9	7.7
WR	Tyrell Williams Oak WR	LAC	W 26-24	8.4	2.5
TE	Kyle Rudolph Min TE	@Dal	W 28-24	5.1	15.4
FLEX	Tyler Lockett Sea WR	@SF	W 27-24	10.7	2.9
FLEX	Demaryius Thomas NYJ WR	NYG	W 34-27	5.0	8.4
D/ST	Jets D/ST NYJ D/ST	NYG	W 34-27	8.0	17.0
K	Harrison Butker KC K	@Ten	L 32-35	8.7	14.0
		TOTALS		104.6	122.5

- Each week you have to predict which players are going to score the most fantasy points
- Calculation of fantasy points can differ by league but usually based on:
  - Yards (passing, rushing, receiving)
  - TDs
  - Turnovers (i.e. Interception or fumble)
  - Defence / Special Team / Kicking statistics

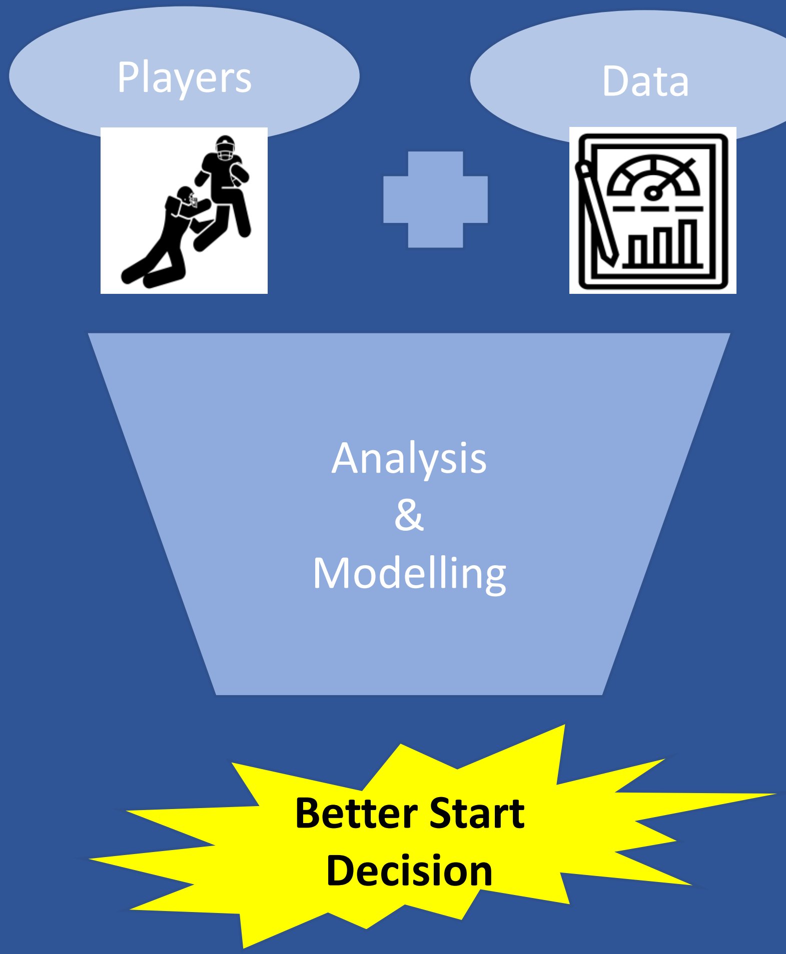
But often it is not clear who the best person to start is



Do I pick the player who is…

- Accumulating the most fantasy points this season? (i.e. consistent)
- After putting up a monster fantasy day last week? (i.e. the hot hand)
- Gaining the most yards?
- Scoring the most touchdowns? (after all, they get me more points)
- Playing the weaker opposition defence this week?
- Playing on the better offense in the league?

Objective is to help fantasy football owners make better decisions on who to start in their fantasy team each week



Dataset

Looked at player performance data between 2015 and 2019 seasons

Player	Player name and position
Player Team & Opposition	Team and coaches for player's team and opposition
Date & Time	Season, gameweek, date and time of game
Game Details	Game location, stadium type, and surface type
Game Score	Game results, points for and against
Player Statistics	Passing, Rushing, and Receiving statistics for each player
Fantasy Points	Fantasy point statistics for each player

Where fantasy points related to a specific league scoring system

How were Fantasy Points calculated?

Players are assigned fantasy points for the following statistics

Yards

Every 25 passing yards = 1 pt
Every 10 rushing yards = 1 pt
Every 10 receiving yards = 1 pt

300-399 passing yards (bonus) = 2 pt
400+ passing yards (bonus) = 3 pt

100-199 rushing yards (bonus) = 2 pt
200+ rushing yards (bonus) = 3 pt

100-199 receiving yards (bonus) = 2 pt
200+ receiving yards (bonus) = 3 pt

Scores

Touchdown (Passing, Rushing or Receiving) = 6 point

2pt Conversion (Passing, Rushing or Receiving) = 2 point

Negative Plays

Interception = -2pt

Sacked = -0.5pts

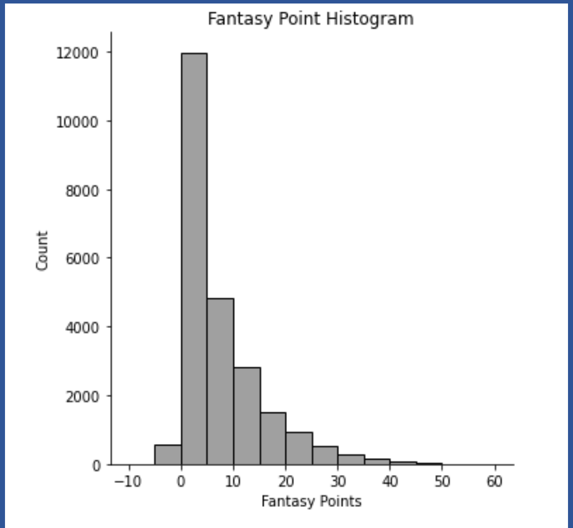
Fumble (Lost) = -2pt

Exploratory Data Analysis

Focus on understanding what influences a player's fantasy performance each week

How many fantasy points do players score?

Histogram of Fantasy Point Distribution

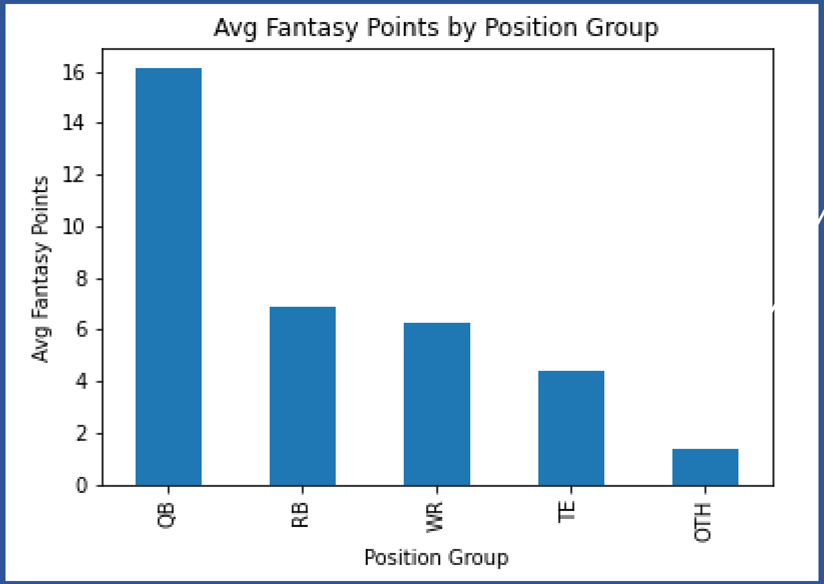


7.2 average fantasy points per player

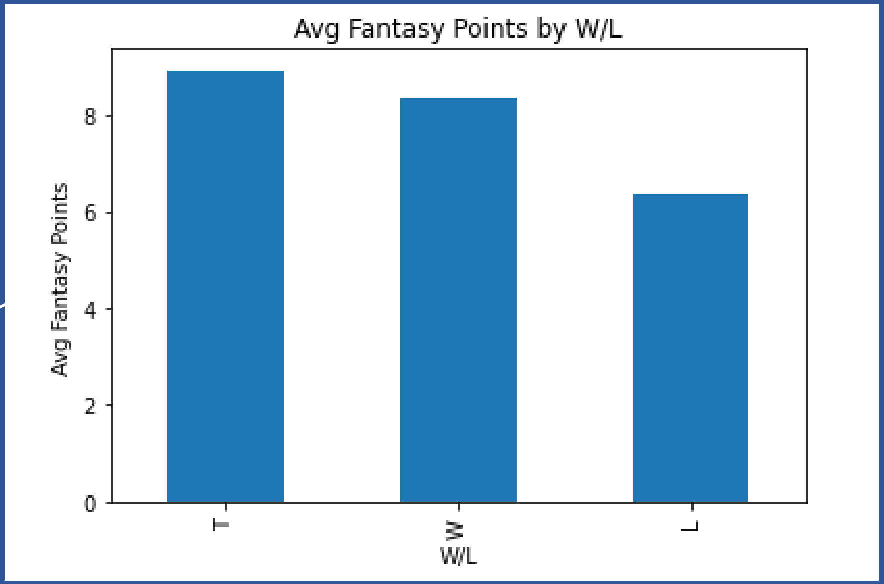
73% of players score 10 points or less

61.3 points is the highest in dataset by QB Drew Brees (Week 8, 2015 vs New York Giants)

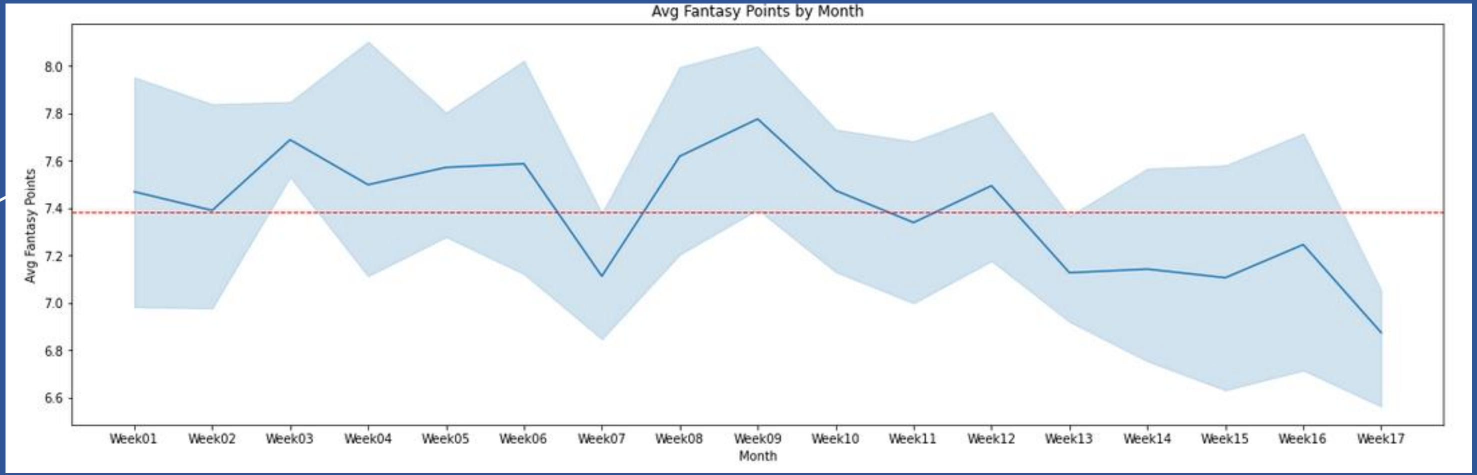
Exploring this data threw up interesting insights



QBs score twice as many points as other positions



On average, players score more in wins & ties than in losses



Fantasy points seem to trend down as season progresses

However, there were issues relating to player and game statistics

Issues

1 Not independent

- Fantasy statistics are calculated from raw player statistics, meaning they are not independent

2 Information not available at decision time

- Fantasy owners don't have access to the score of a game or how player performs until after the game – when it is too late

Solution

Convert player, score and result statistics into rolling averages for previous 4-weeks

PLAYER_NAME	TEAM	OPP	DATE	W/L	PTS_FOR	PTS_AGT	PASSYDS	PASSTD	INT	RUSHYDS	RUSHTD	FUMLST	REC	RECYDS	RECTD	FAN_TOT
Danny Amendola	Patriots	Steelers	2015-09-10	W	28	21	0	0	0	0	0	0	2	24	0	2.4
DeAngelo Williams	Steelers	Patriots	2015-09-10	L	21	28	0	0	0	127	0	0	1	5	0	15.2
Dion Lewis	Patriots	Steelers	2015-09-10	W	28	21	0	0	0	69	0	0	4	51	0	12.0
Rob Gronkowski	Patriots	Steelers	2015-09-10	W	28	21	0	0	0	0	0	0	5	94	3	27.4
Markus Wheaton	Steelers	Patriots	2015-09-10	L	21	28	0	0	0	0	0	0	3	55	0	5.5

Which were rectified by calculating 4-week average for all these statistics

PLAYER_NAME	DATE	FAN_TOT	FAN_4WK_AVG	FAN_4WK_AVG_SHIFT
Calvin Ridley	2018-09-16	12.7	12.700	NaN
Calvin Ridley	2018-09-23	35.5	24.100	12.700
Calvin Ridley	2018-09-30	17.5	21.900	24.100
Calvin Ridley	2018-10-07	3.8	17.375	21.900
Calvin Ridley	2018-10-14	4.7	15.375	17.375
Calvin Ridley	2018-10-22	4.3	7.575	15.375
Calvin Ridley	2018-11-04	13.7	6.625	7.575
Calvin Ridley	2018-11-11	4.0	6.675	6.625

Steps to calculate

1 Fantasy Points scored in a week

2 Average fantasy points scored in last 4 weeks (including this week)

3 Average fantasy points scored in previous 4 weeks (not including this week)

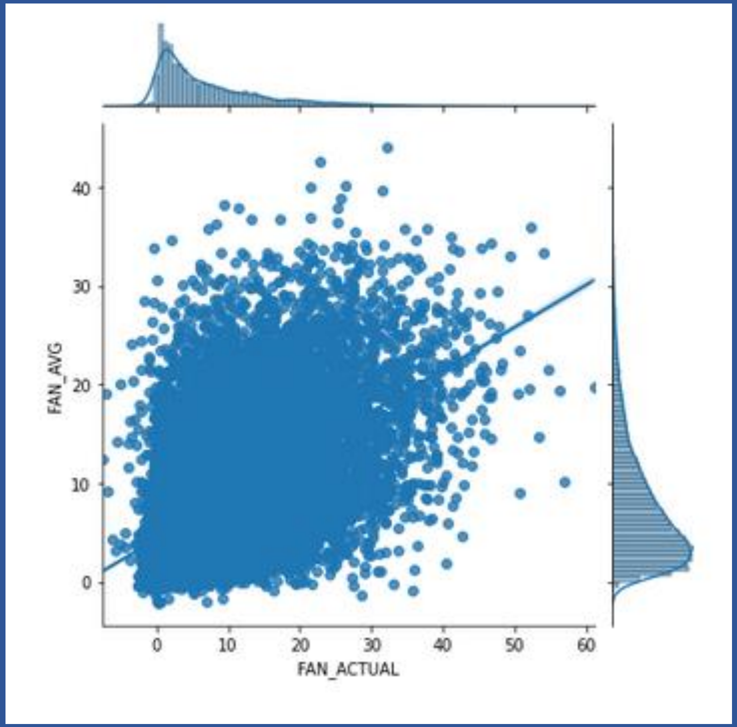
4 Drop values calculated in Step 2

5 Calculate 4-week rolling averages with a shift for all player and game statistics (and drop actual and 4 week average used to calculate)

6 The process of creating 4-week rolling averages created 1,107 rows with missing data – these were dropped from dataset*

PLAYER	TEAM	OPP	DATE	WIN/TIE_AVG	PTS_FOR_AVG	PTS_AGT_AVG	PASSYDS_AVG	PASSTD_AVG	INT_AVG	RUSHYDS_AVG	RUSHTD_AVG	REC_AVG	RECYDS_AVG	RECTD_AVG	FAN_AVG	FAN_ACTUAL
Demaryius Thomas	Broncos	Chiefs	2017-10-30	0.25	10.50	20.00	0.0	0.0	0.0	0.0	0.0	4.75	62.75	0.00	6.275	6.6
Allen Lazard	Packers	Raiders	2019-10-20	0.75	27.75	24.00	0.0	0.0	0.0	0.0	0.0	2.50	36.00	0.50	6.600	4.2
C.J. Uzomah	Bengals	Colts	2018-09-09	0.50	17.75	27.75	0.0	0.0	0.0	0.0	0.0	1.75	13.25	0.25	2.825	0.4
Seth Roberts	Raiders	Broncos	2016-11-06	0.75	26.75	24.25	0.0	0.0	0.0	0.0	0.0	3.00	40.50	0.25	5.550	3.2
Chris Hogan	Patriots	Chiefs	2018-10-14	0.50	26.50	22.00	0.0	0.0	0.0	0.0	0.0	2.50	33.00	0.50	6.300	7.8

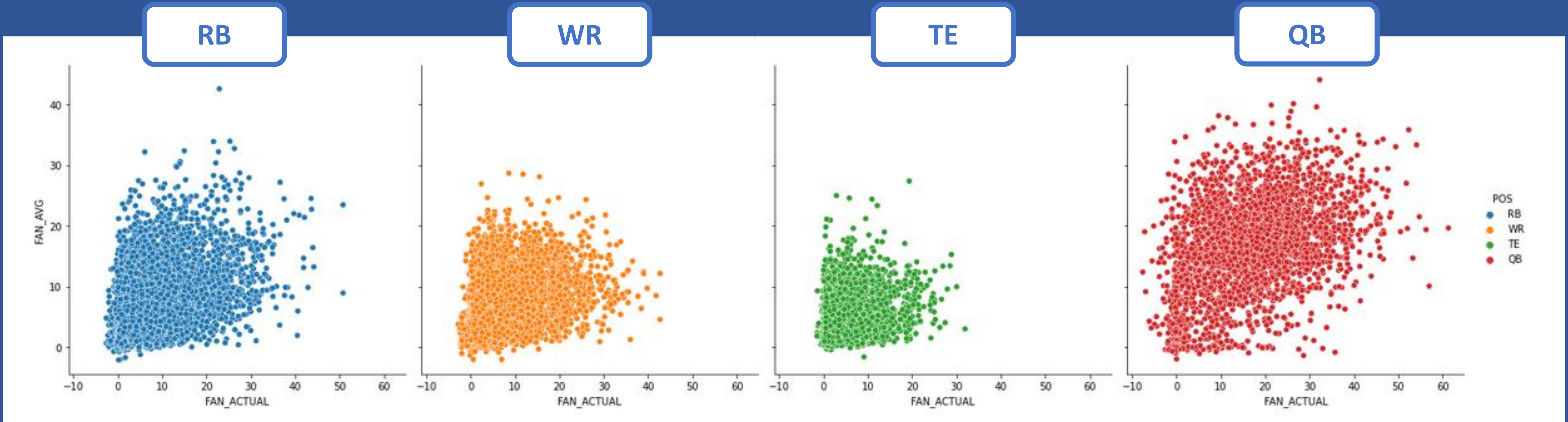
Found that fantasy performance is significantly correlated with fantasy performance over the previous 4 weeks



0.56 pearson coefficient

Statistically significant after running permutation test for higher correlation coefficient and achieving a 0.0 p-value

While the correlation of other statistics varies by position



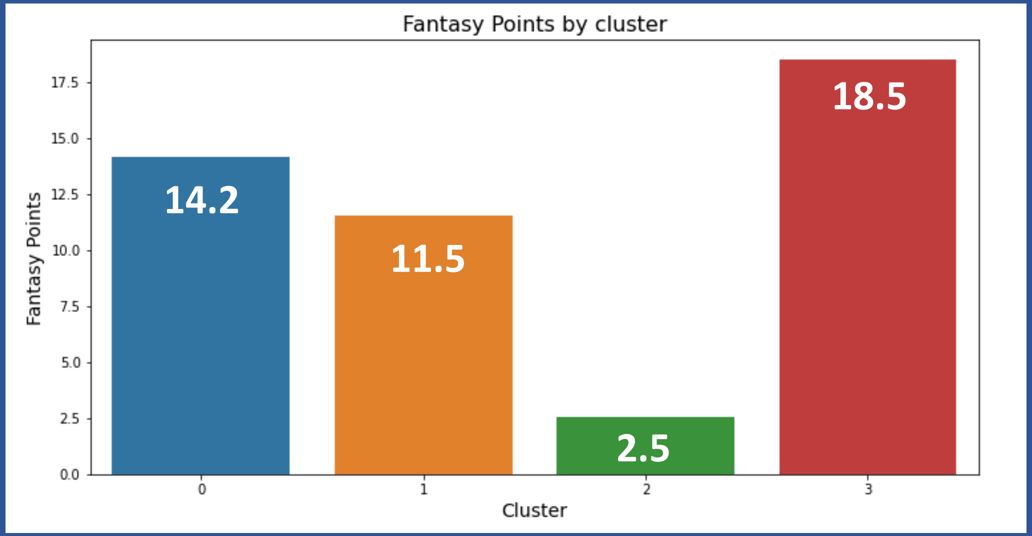
RB	WR	TE	QB
0.48 fantasy points	**0.35** fantasy points	**0.32** fantasy points	**0.43** fantasy points
0.47 rushing attempts	**0.38** targets	**0.37** targets	**0.42** pass yards
0.40 rushing yards	**0.37** receiving yards	**0.37** receiving yards	**0.40** pass completions

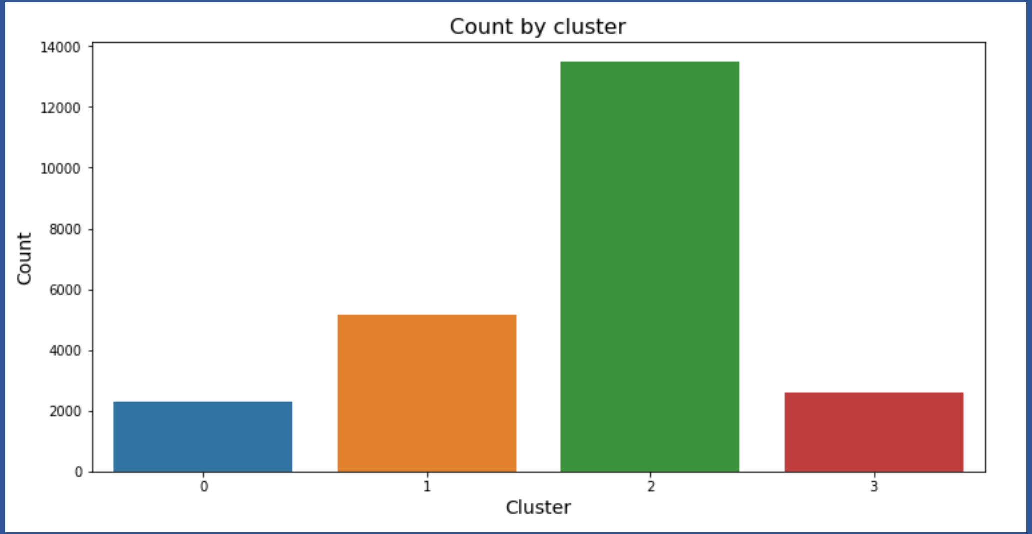
Using K-means analysis we could identify four interesting clusters in the data

Four distinct groups of players

- Three groups are fantasy relevant players – split by the type of statistics they perform best at
- One group is not fantasy relevant but consists of over half the player in dataset

Cluster	Scoring	Count	Prominent Position	Main Statistics
0	2nd	4th	RB	Rushing
1	3rd	2nd	WR/TE	Receiving
2	4th	1st	All	None
3	1st	3rd	QB	Passing





Feature Selection

EDA identified columns and rows to remove or transform in our dataset

Drop Columns

- Player Name column
- All team and coaches columns
- Week and date columns
- Game statistic columns (replaced with 4-week rolling averages)
- Player statistic columns (replaced with 4-week rolling averages)

Create Binary & Dummy Columns

- Transform Home/Away, Surface, Stadium, Day and Result into binary columns (i.e. 1 if True, 0 if False)
- Create dummy variables columns for Season, Month and Position features

Drop Rows

- NaN values
- ‘Other’ position group

Why?

Remove high dimension columns where proxy information available

Replace columns with information that reflects that available to fantasy owner at decision time

Transform to binary and dummy columns to support modelling

Remove rows that will impact on modelling

This means left with 2 targets and 48 features to split & scale for before modelling

2 targets*

- FAN_ACTUAL
- CLUSTER

48 features

- FAN_AVG
- PASSCOMP_AVG
- PASSATT_AVG
- PASSCOMP%_AVG
- PASSYDS_AVG
- PASSTD_AVG
- INT_AVG
- QBRAT_AVG
- SACK_AVG
- SACKYDS_AVG
- PASSYDS_300_AVG
- PASSYDS_400_AVG
- RUSHATT_AVG
- RUSHYDS_AVG
- RUSHTD_AVG
- FUM_AVG
- FUMLST_AVG
- RUSHYDS_100_AVG
- RUSHYDS_200_AVG
- TGTS_AVG
- REC_AVG
- RECYDS_AVG
- RECTD_AVG
- RECYDS_100_AVG
- RECYDS_200_AVG
- PTS_FOR_AVG
- PTS_AGT_AVG
- WIN/TIE_AVG
- OPP_PTS_FOR_AVG
- OPP_PTS_AGT_AVG
- OPP_WIN/TIE_AVG

- HOME
- DOME
- GRASS
- SUNDAY
- POS_RB
- POS_TE
- POS_WR
- SEASON_2016
- SEASON_2017
- SEASON_2018
- SEASON_2019
- MONTH_January
- MONTH_November
- MONTH_October
- MONTH_September
- TIME_Night
- TIME_Noon

Train / Test Split

Y = FAN_ACTUAL
X = Features

75% / 25% split – Training to Test

Scaling

- Standardization applied to all continuous variables (left-side of features list)
- Dummy and binary variables are not scaled (right-side features list)

*** Two potential target variables for our model. Initial focus is on 'FAN_ACTUAL' and building a model to predict the number of fantasy points.**

Variance, covariance and recursive feature elimination used to further reduce features

Eliminated Features

Method	Eliminated Features	Count
Variance Feature Elimination *Eliminate all columns where the 90%+ of the variables are similar*	MONTH_January	1
Covariance Feature Elimination *Eliminate columns which have 0.8 or higher correlation (or Pearson coefficient)*	'PASSATT_AVG', 'PASSCOMP%_AVG', 'PASSYDS_AVG', 'PASSTD_AVG', 'INT_AVG', 'QBRAT_AVG', 'SACK_AVG', 'SACKYDS_AVG', 'RUSHYDS_AVG', 'REC_AVG', 'RECYDS_AVG'	11
Recursive Feature Elimination *Recursively run estimator (SVR with 'linear' kernel) over features to rank and prune features to attain optimal number of features for dataset*	'PASSYDS_400_AVG', 'FUM_AVG', 'RUSHYDS_100_AVG', 'RUSHYDS_200_AVG', 'RECYDS_100_AVG', 'RECYDS_200_AVG', 'PTS_FOR_AVG', 'PTS_AGT_AVG', 'OPP_PTS_FOR_AVG', 'OPP_WIN/TIE_AVG', 'POS_RB', 'SEASON_2016', 'SEASON_2017', 'SEASON_2018', 'SEASON_2019', 'MONTH_November', 'MONTH_October'	17

TOTAL ELIMINATED **29**

For our final dataset with 19 features and our target variable

1 target

- FAN_ACTUAL

19 features

- FAN_AVG
- PASSCOMP_AVG
- PASSYDS_300_AVG
- RUSHATT_AVG
- RUSHTD_AVG
- FUMLST_AVG
- TGTS_AVG
- RECTD_AVG
- WIN/TIE_AVG
- OPP_PTS_AGT_AVG
- HOME
- DOME
- GRASS
- SUNDAY
- POS_TE
- POS_WR
- MONTH_September
- TIME_Night
- TIME_Noon

Modelling

Four regression models were chosen for machine learning

1	**Linear Regression**	Ordinary least squares (OLS) regression is a linear model that is fitted by minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable configured as a straight line.
2	**SGD Regressor**	Stochastic Gradient Descent (SGD) Regressor is a linear model that is fitted by minimizing a regularized empirical loss with SGD - the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing strength schedule (aka learning rate).
3	**ElasticNet**	ElasticNet is a linear regression model trained with both l1 and l2 -norm regularization of the coefficients. This combination allows for learning a sparse model where few of the weights are non-zero like Lasso, while still maintaining the regularization properties of Ridge. Elastic-net is useful when there are multiple features which are correlated with one another.
4	**Random Forest Regressor**	A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

*** Regression models used as our target variable is continuous**

Each model had hyperparameters tuned using cross-validation

1	**Linear Regression**	✘	
2	**SGD Regressor**	loss = ‘squared_loss’ penalty = ‘adaptive’ alpha = 0.01 learning_rate = ‘l1’ Eta0 = 0.01	**RandomizedSearchCV**
3	**ElasticNet**	l1_ratio = 1 Alpha = 0.01	**ElasticNetCV**
4	**Random Forest Regressor**	n_estimator = 1,000 max_depth = 10 max_features = ‘sqrt’	**GridSearchCV**

Each model was evaluated with three metrics

1	**Root mean squared error (RMSE)**	Sample standard deviation of difference predicted values and observed values
2	**Mean absolute error (MAE)**	Average absolute difference between the predicted values and the observed values
3	**R-Squared (R2)**	Explains how well features in model explain the variability in our target variable (fantasy points)

The models were fitted and evaluated on training data

Model	RMSE	MAE	R-Squared
Linear	6.39	4.65	0.364
SGD Regressor	6.39	4.66	0.363
ElasticNet	6.39	4.65	0.363
Random Forest	5.66	4.24	0.501

Random Forest was the best model when applied to all training data

- Lowest RMSE
- Lowest MAE
- Highest R-squared

As well as evaluated with cross-validation to understand how performance would generalise

Model	RMSE	MAE	R-Squared
Linear	6.40	4.66	0.361
SGD Regressor	6.41	4.66	0.358
ElasticNet	6.68	5.03	0.304
Random Forest	6.65	4.90	0.310

Linear was the best model when 5-fold cross-validation was applied to the training data

- Lowest RMSE
- Lowest MAE
- Highest R-squared

Finally, models were evaluated on our test data

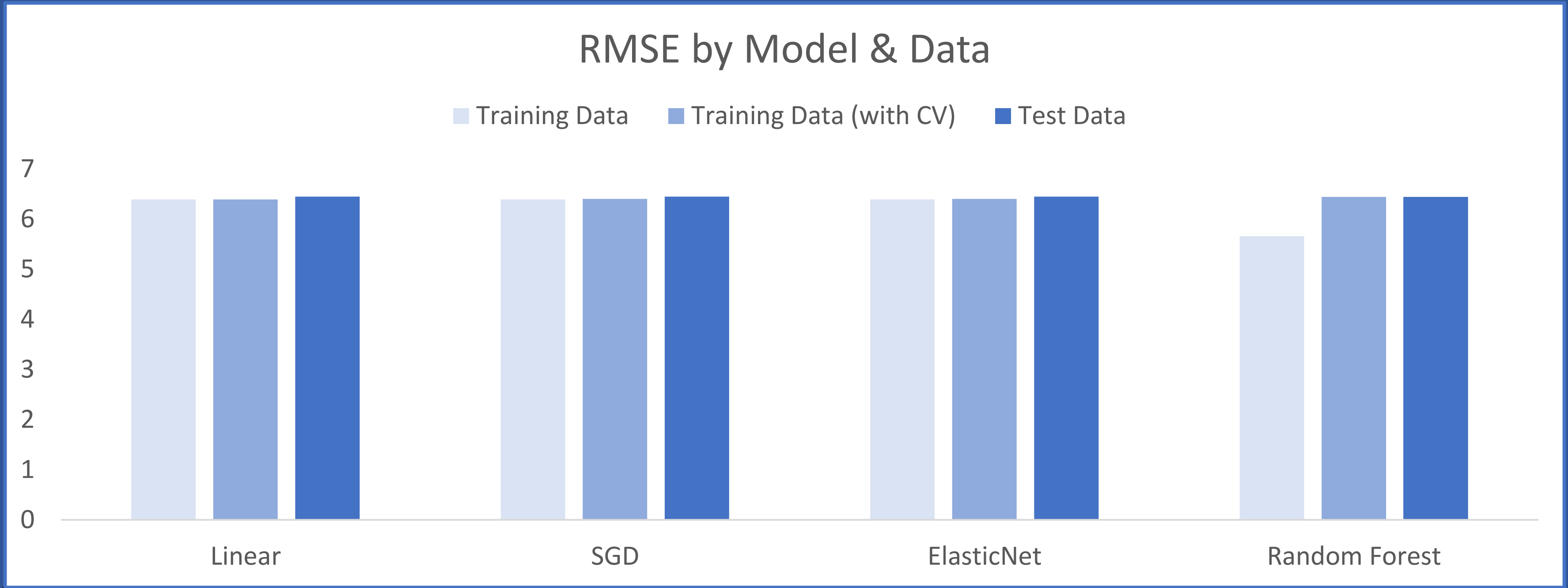
Model	RMSE	MAE	R-Squared
Linear	6.45	4.69	0.350
SGD Regressor	6.45	4.69	0.350
ElasticNet	6.45	4.69	0.350
Random Forest	6.44	4.71	0.350

Very little to differentiate models when applied to test data

- Random Forest model performed best in relation to RMSE
- Other models performed better in relation to MAE
- R-squared consistent across all models

- Preference depends on whether RMSE or MAE is better evaluation metric for the data
- RMSE considered better metric for this data set – ability to penalise higher differences

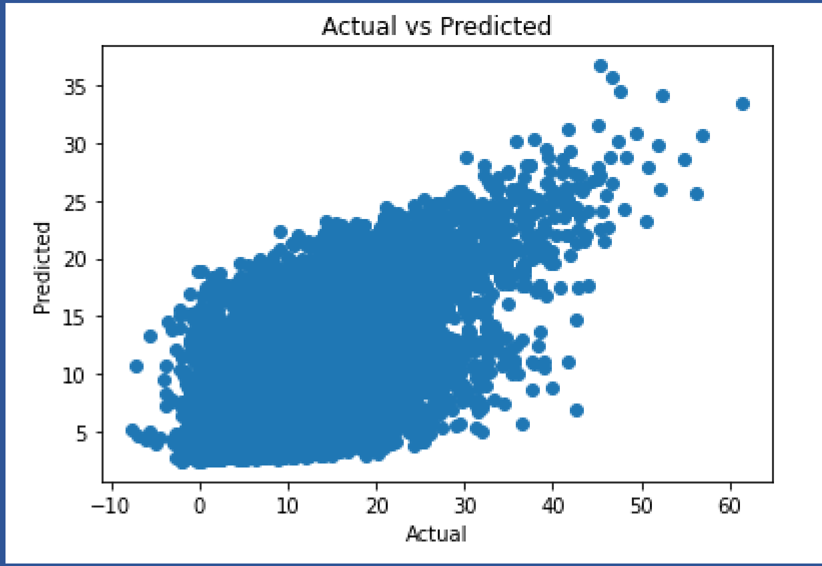
Random Forest model proves to be the best model when focusing on RMSE

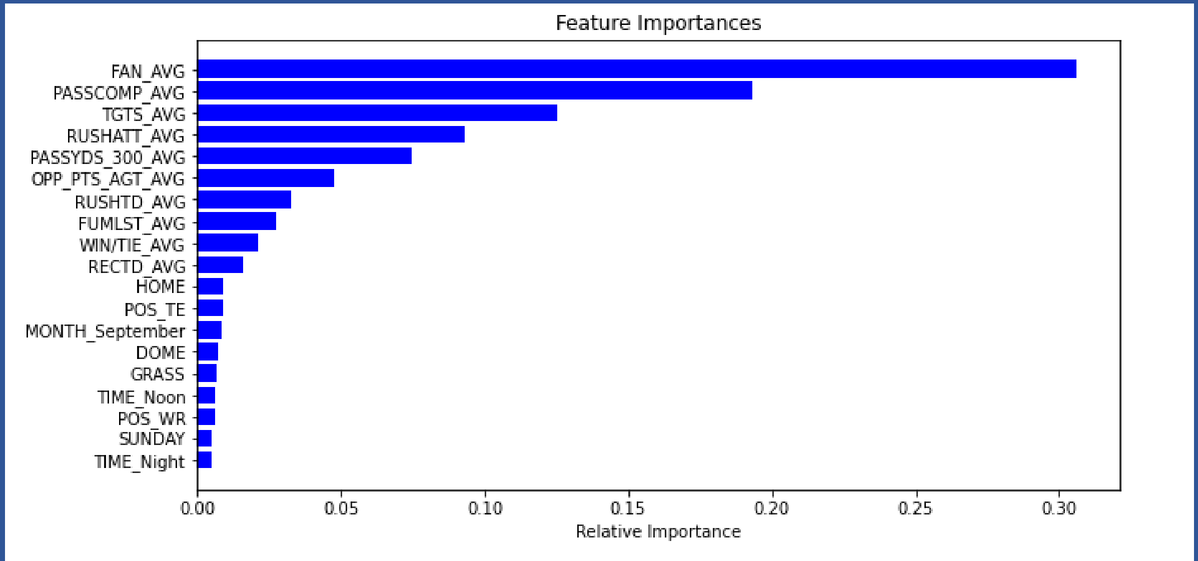


	Linear	SGD	ElasticNet	Random Forest
Training	6.39	6.39	6.39	5.66
Training (with CV)*	6.39	6.40	6.40	6.44
Test	6.45	6.45	6.45	6.44

*** Applied 5-fold cross-validation (using cross_val_score) on training set to judge how well results would generalise**

Fantasy performance over previous weeks proved to be the most important feature in model





4.71 fantasy points is the average prediction error of this model

4-week fantasy performance is most important feature

- Pass completion, Targets, and Rushing attempts are the next most important features – make sense given they are indicators for each of the key positions on the field (QB, RB, WR/TE)

Conclusion

Who did you pick?

?

?

?

?

?

?

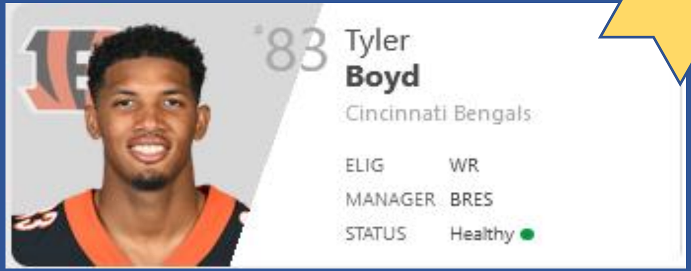
Our model suggested picking Adam Thielen, Tyler Boyd, and Marquez Valdes-Scantling

6.2

4.7

9.3

6.9



8.0

6.2

In real life, Corey Davis overperformed Tyler Boyd but are other two players were the right choice

1.0

6.0

8.5

21.3

3.5

14.0

I did not make the right choice

Model would not have selected the optimal team…

…but using it would've gotten me the win!

Colour Legend

- Selection same as Model and best in reality
- Selection same as Model but wrong in reality
- Selection opposite of Model and wrong in reality

MY TEAM 118.2 123.6 WINNING TEAM

Overview | Full Box Score | Scoring Breakdown

Bresburgh Steelers Box Score

SLOT	PLAYER	OPP	STATUS	PROJ	FPTS
QB	Russell Wilson Sea QB	Bal	L 16-30	25.9	15.8
RB	Saquon Barkley NYG RB	Ari	L 21-27	18.6	14.0
RB	Frank Gore Buf RB	Mia	W 31-21	12.0	6.6
WR	Adam Thielen Min WR	@Det	W 42-30	10.1	8.5
WR	Tyler Boyd Cin WR	Jax	L 17-27	8.9	3.5
TE	George Kittle SF TE	@Wsh	W 9-0	9.6	3.8
FLEX	Latavius Murray NO RB	@Chi	W 36-25	8.6	29.0
FLEX	Emmanuel Sanders SF WR	@Wsh	W 9-0	7.4	6.0
D/ST	Chiefs D/ST KC D/ST	@Den	W 30-6	7.0	25.0
K	Mason Crosby GB K	Oak	W 42-24	7.8	6.0
		TOTALS		115.8	118.2
Bench	Robby Anderson NYJ WR	NE	L 0-33	5.6	1.0
Bench	Corey Davis Ten WR	LAC	W 23-20	6.1	14.0
Bench	Marquez Valdes-Scantling GB WR	Oak	W 42-24	7.3	21.3
Bench	Chase Edmonds Ari RB	@NYG	W 27-21	4.9	35.0
Bench	Tony Pollard Dal RB	Phi	W 37-10	3.1	2.8

Steattle Stehawks Box Score

SLOT	PLAYER	OPP	STATUS	PROJ	FPTS
QB	Derek Carr Oak QB	@GB	L 24-42	15.1	19.6
RB	Ezekiel Elliott Dal RB	Phi	W 37-10	17.1	22.7
RB	Derrick Henry Ten RB	LAC	W 23-20	15.3	16.8
WR	Keenan Allen LAC WR	@Ten	L 20-23	9.3	6.1
WR	Calvin Ridley Atl WR	LAR	L 10-37	9.1	3.0
TE	Jason Witten Dal TE	Phi	W 37-10	4.6	3.3
FLEX	Kenny Stills Hou WR	@Ind	L 23-30	5.1	12.5
FLEX	Phillip Dorsett II NE WR	@NYJ	W 33-0	8.0	10.6
D/ST	Jaguars D/ST Jax D/ST	@Cin	W 27-17	7.8	21.0
K	Wil Lutz NO K	@Chi	W 36-25	5.9	8.0
		TOTALS		97.3	123.6
Bench	Josh Gordon Sea WR	Bal	L 16-30	0.0	0.0
Bench	Baker Mayfield Cle QB	*BYE*	--	--	0.0
Bench	Vance McDonald Pit TE	*BYE*	--	--	0.0
Bench	Peyton Barber TB RB	*BYE*	--	--	0.0
Bench	Randall Cobb Dal WR	Phi	W 37-10	6.3	2.9

Conclusions

Good Start

- As example shows, this model can be a useful aid to help fantasy owners make decisions on who to select in a given game week
- Predictions will error on average by +/- 4.7 fantasy points using the model
- Often fantasy owners need to pick between players and not guess precise numbers
- Model also helps our understanding of fantasy performance and that previous 4-week performance across fantasy performance and other player statistics are useful indicators

Room to Improve

- One point can make a huge difference in fantasy football (see previous slide)
- Aim would be to reduce error term to make more accurate predictions
- This would give fantasy football owners more comfort they are making the right decision

Next Steps

How to Improve

Develop rolling averages further

- Our model used 4-weeks weighted equally but alternatives should be used and tested

Better use of clusters

- Clusters were identified in EDA but not worked into the modelling. In future could look to use as target variable or feature

Model by Position Group

- EDA identified that performance differs significantly by position group so future models could be built to focus on one position for more accurate results

New Data

- Look to include more seasons of data or look for new data sources and information to contribute to the model (i.e. Offensive Line performance or Player Salary)

Different Feature Engineering Models

- Look to use different models to feature engineer like Principal Component Analysis

New Technique to handle missing values

- We removed missing values from the dataset but could to replace these values with appropriate estimate

Thank You