Predicting NFL Defensive Player Longevity

Billie Kim
Casey Ng
Ethan Panal

Problem & Stakeholders

- Career longevity for football players is very volatile
- NFL General Managers (GMs) manage multi-million dollar budgets
- **Solution**: Aid GMs with roster strategy through career length predictions for defensive players

NFL Team Salary Cap Tracker

A real-time look at the 2023 **salary cap totals** for each **NFL team**, including estimated cap space. Assumes a $224,800,000 team salary cap.

Cap Tracker | Cash Tracker | Combined AAV | Acquired By

2023 | $ | UPDATE

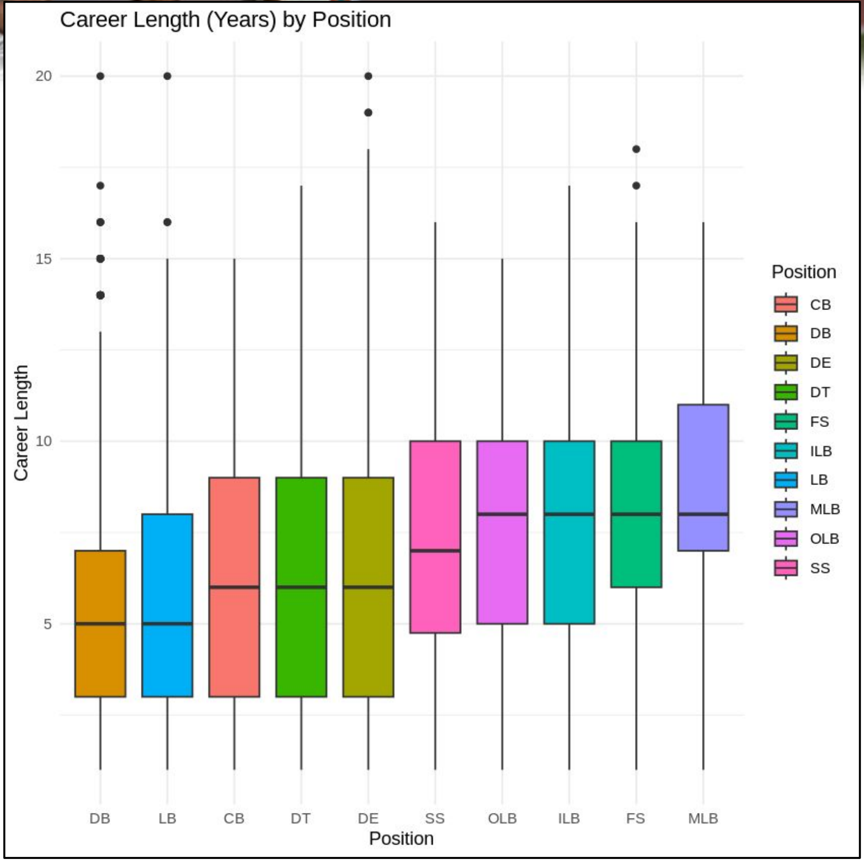
RANK	TEAM	WIN %	SIGNED	AVG AGE	ACTIVE	DEAD	TOTAL CAP	CAP SPACE (ALL)
1	San Francisco 49ers	1.000	53	26.85	$157,880,314	$27,749,981	$193,136,441	$44,383,827
2	Cleveland Browns	1.000	53	26.4	$177,026,840	$18,736,950	$214,569,459	$37,202,375
3	Dallas Cowboys	1.000	53	26.13	$192,850,638	$17,245,597	$218,991,314	$13,912,537
4	Arizona Cardinals	0.000	53	26.3	$122,487,146	$47,504,518	$215,684,781	$13,614,760
5	Cincinnati Bengals	0.000	53	25.75	$205,372,108	$2,804,316	$213,194,157	$13,414,579
6	Las Vegas Raiders	1.000	53	26.91	$180,549,657	$34,193,561	$223,651,114	$11,350,154
7	Tennessee Titans	0.000	53	26.51	$171,168,908	$39,208,060	$219,878,487	$11,020,547

Dataset

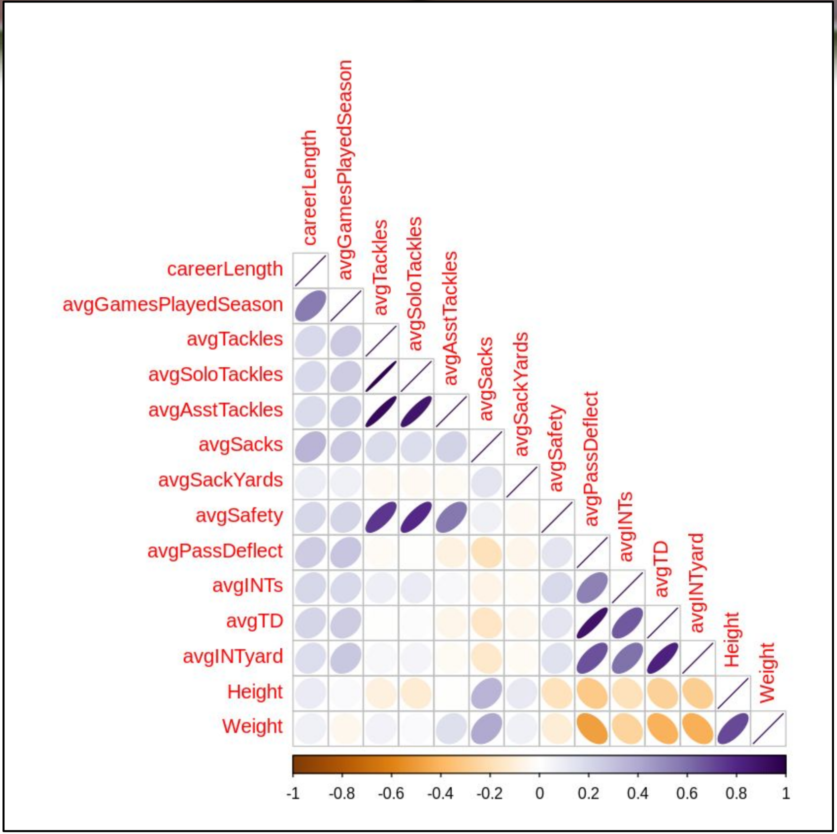
- 6,000+ retired defensive player stats from 1982-2019
- Average performance + physical features

Player_Id	careerLength	avgGamesPlayedSeason	avgTackles	avgSoloTackles	avgAsstTackles	avgSacks	avgSackYards	avgSafety	avgPassDeflect	avgINTs	avgTD	avgINTyard	Position	Height	Weight
<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>
a-j-duhe	8	13.50000	0.00000	0.00000	0.000000	1.0625000	0.125	0.000000	0.3750000	0.0000000	2.25000	2.25000	LB	76	247
a-j-edds	2	8.50000	6.50000	1.50000	5.000000	0.0000000	0.000	0.000000	0.0000000	0.0000000	0.00000	0.00000	LB	76	256
a-j-francis	4	2.25000	4.25000	0.75000	3.500000	0.0000000	0.000	0.000000	0.0000000	0.0000000	0.00000	0.00000	DT	77	332
a-j-hawk	11	14.45455	74.72727	47.27273	27.454545	1.8181818	0.000	3.000000	0.8181818	0.0000000	11.00000	5.80000	LB	73	240
a-j-jefferson	4	10.75000	26.50000	24.00000	2.500000	0.0000000	0.000	4.750000	0.5000000	0.0000000	0.25000	0.25000	CB	73	190
a-j-jenkins-2	2	10.50000	0.00000	0.00000	0.000000	1.0000000	0.000	0.000000	0.0000000	0.0000000	0.00000	0.00000	LB	74	237
a-j-johnson-2	7	10.14286	0.00000	0.00000	0.000000	0.1428571	0.000	0.000000	1.2857143	0.2857143	28.71429	15.02857	DB	68	175
a-j-schable	1	11.00000	10.00000	8.00000	2.000000	0.0000000	0.000	0.000000	0.0000000	0.0000000	0.00000	0.00000	DE	75	281
a-j-tarpley	1	14.00000	7.00000	4.00000	3.000000	1.0000000	0.000	2.000000	2.0000000	0.0000000	40.00000	20.00000	OLB	72	232
aaron-beasley	9	13.44444	17.66667	14.55556	3.111111	0.9444444	0.000	3.888889	2.6666667	0.2222222	54.11111	17.06667	CB	72	205

Data Exploration



- Median careers were 5-8 years position dependent



- Avg Games Played per Season was highest correlated with player's career length

Regression Model(s)

- Model results: *Simple Linear Regression*

Best Model: Games Played Model (LR1)

'Predicted mean career length = 0.00736831058447645 + 0.547824698735212 * (average games played)'

```
Call:
lm(formula = careerLength ~ avgGamesPlayedSeason, data = df_numeric)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.7726 -1.9100 -0.1563  1.8274 12.6518 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          0.007368   0.115955   0.064    0.949    
avgGamesPlayedSeason 0.547825   0.010167  53.882   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.939 on 6126 degrees of freedom
Multiple R-squared:  0.3215,	Adjusted R-squared:  0.3214 
F-statistic:  2903 on 1 and 6126 DF,  p-value: < 2.2e-16
```

	R2	RSE
LR1	0.32140	2.939
LR2	0.12330	3.340
LR3	0.06262	3.454
LR4	0.04801	3.481
LR5	0.04366	3.489

Regression Model(s)

- Model results: *Multiple Linear Regression*

Best Model: Model 5 (MR5)

```
Call:
lm(formula = careerLength ~ avgGamesPlayedSeason + avgSacks +
    avgPassDeflect + avgTD + avgSafety + avgINTs, data = df_numeric)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6140 -1.7538 -0.1782  1.6176 13.6881

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.406001   0.110092   3.688 0.000228 ***
avgGamesPlayedSeason  0.418445   0.010714  39.055  < 2e-16 ***
avgSacks              0.682480   0.027811  24.540  < 2e-16 ***
avgPassDeflect        0.794205   0.081923   9.695  < 2e-16 ***
avgTD                -0.022583   0.005563  -4.060 4.97e-05 ***
avgSafety             0.159117   0.023902   6.657 3.04e-11 ***
avgINTs               2.011021   0.406882   4.943 7.92e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.765 on 6121 degrees of freedom
Multiple R-squared:  0.3998,    Adjusted R-squared:  0.3992
F-statistic: 679.4 on 6 and 6121 DF,  p-value: < 2.2e-16
```

	R2	RSE
MR1	0.3306	2.919
MR2	0.3721	2.827
MR3	0.3535	2.869
MR4	0.3969	2.771
MR5	0.3992	2.765

Takeaways

- Use model or not?
  - **Why or why not?**

- Models are not ready to be deployed
- Generally desire higher accuracy scores to increase prediction power and decrease prediction error
- Additional data preprocessing and cleaning may help enhance model performance

However…

- Our Models can still prove to be useful to make soft estimates that can still provide useful insights for General/Team Managers

Next Steps

- More optimized model and feature selection methods
- Using more advanced machine learning techniques
- Apply to the offense