

Predicting rank of players in NFL

Wendy Yang

January 17, 2018

Data Science Part Time Course

Why do we care?

Fantasy Football Market is estimated at over \$7B per year

Personally took 11th place in my fantasy league this year
☹️

The screenshot shows a mobile app interface for Fantasy Football. At the top, the status bar displays the time as 2:25. The app header reads "Season Story | Fantasy Football | Ya..." with a close button (X). Below the header, there's a circular profile picture of a cityscape and a large yellow circle containing a "C-" grade, labeled "Season Grade". To the left of the grade, the team name "The golden cloaks" is shown with a dropdown arrow. Below the team name, statistics are listed: "11TH PLACE 6-8-0" and "TOTAL POINTS 988.68". Further down, "POINTS/WEEK 70.62" is displayed. A horizontal blue line separates this section from the "Season Story" section below. Under "Season Story", there are four links: "Season Start: Draft Strategies", "In-Season Management", "Projected and Final Standings", and "Season Grade Summary". The "Season Start: Draft Strategies" link is highlighted with a green circle containing a "B-". Below this link, a paragraph explains that the chart below illustrates how positions were valued in the draft by each team, measuring the weighted distribution of draft picks. It notes that positions drafted earlier and more often are valued more highly, and that the height of the bar in the chart represents the value of a position. It concludes by asking the user what position they valued most. At the bottom of the screen, a black navigation bar contains three icons: a back arrow, a home circle, and a recent apps square. Above this bar, the text "Draft Value by Position" is visible, indicating the title of the chart that was partially cut off.

Season Story | Fantasy Football | Ya... X

The golden cloaks ▾

11TH PLACE 6-8-0 TOTAL POINTS 988.68

POINTS/WEEK 70.62

Season Grade

Season Story

[Season Start: Draft Strategies](#) [In-Season Management](#)

[Projected and Final Standings](#) [Season Grade Summary](#)

B- **Season Start: Draft Strategies**

Your season kicked off with the draft. The chart below illustrates how positions were valued in the draft by each team. It measures the weighted distribution of draft picks. Positions that were drafted earlier and more often are valued more highly. The taller the bar in the chart the more a position is valued. You can mouse over the bar to see what percent of draft resources was used on each position. What position did you value most?

Draft Value by Position

How do we measure player rank?

This is a continuous variable

Assumptions

- Looking at offense only

- Looking at ending stats across entire season

Where is data from?

<https://github.com/BurntSushi/nflgame>

This pulls JSON files from NFL API

I pulled season stats from 2010 and 2016

Data Cleaning

- had to file a bug, someone removed set[] and changed it to a dictionary which does not have the feature “adds”
- not convinced I actually pulled total season stats, still playing in data

Data exploration

What are the features?

Passing yards, rushing yards, touchdowns

Which features are most useful?

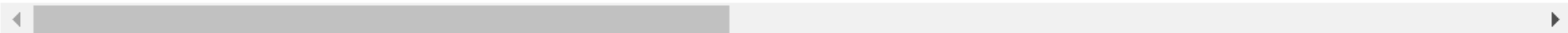
Still determining this

```
In [1]: # read the data into a DataFrame named nfl16
import pandas as pd
path = '../project/'
url = path + 'season2016.csv'
nfl16 = pd.read_csv(url, index_col = 0)
nfl16.head()
```

Out[1]:

	id	home	team	pos	defense_ast	defense_ffum	defense_int	defense_sk	defense_tkl	fumbles_lost	...	re
name												
T.Siemian	00-0032156	no	DEN	QB	NaN	NaN	NaN	NaN	NaN	2.0	...	
A.Janovich	00-0032956	no	DEN	FB	NaN	NaN	NaN	NaN	NaN	NaN	...	
D.Booker	00-0032972	no	DEN	RB	NaN	NaN	NaN	NaN	NaN	3.0	...	
C.Anderson	00-0029854	no	DEN	RB	NaN	NaN	NaN	NaN	NaN	NaN	...	
V.Green	00-0028142	no	DEN	TE	NaN	NaN	NaN	NaN	NaN	NaN	...	

5 rows × 59 columns





```
In [3]: # let's just look at offensive players  
# this includes QB, RB, WR, TE, K  
nfl16.pos.value_counts()
```

```
Out[3]: WR      188  
        DE      143  
        RB      129  
        CB      128  
        TE      103  
        OLB      92  
        DT       89  
        DB       78  
        QB       66  
        LB       65  
        SS       50  
        FS       43  
        ILB      43  
        K        33  
        P        33  
        NT       25  
        OT       17  
        MLB      16  
        C        15  
        FB       15  
        OG       12  
        G         7  
        T         6  
        LS         5  
        SAF         3  
        Name: pos, dtype: int64
```

```
In [4]: # create new DataFrame using only QB, LB, RB, TE, K
```

What's next?

Must figure out best features to use

- Can do this using bar graphs or scatterplots for macro level

- Will try KNN like in NBA exercise – this probably won't work since that was for position rather than rank

- May try regressions instead

Was also looking around to try double layered neural networks but no idea how to explore