Predicting rank of players in NFL

Wendy Yang

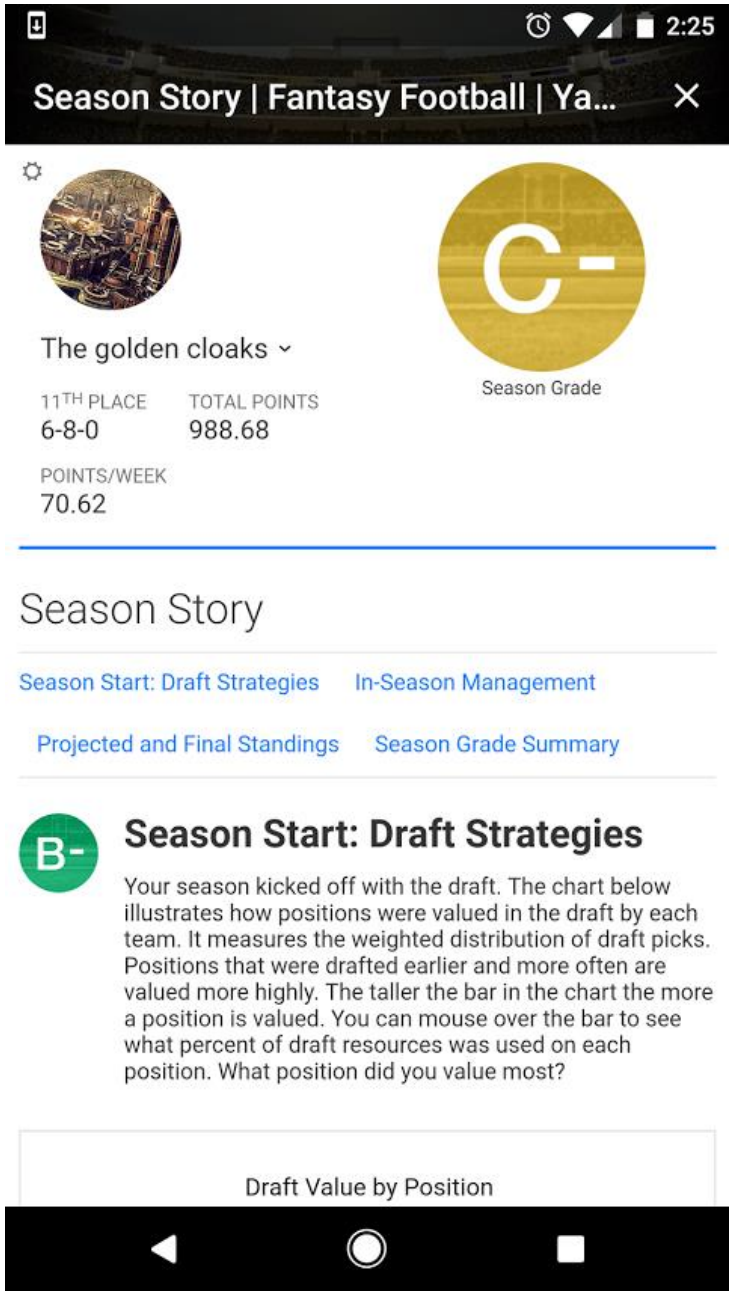
January 17, 2018

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 ☹



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]:

name	id	home	team	pos	defense_ast	defense_ffum	defense_int	defense_sk	defense_tkl	fumbles_lost	...	re
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
```

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