

NBA Draft (College) Basketball Prediction

418 Final Project

Yi Chen

Web Scraping with python to extract
10 years data
from 2009 -2018
via basketball reference. com website

HTML table includes
of games, minutes played, points,
total rebounds and assists

Cleaning the data and get rid of
missing values

Follow the similar steps and do it
multiple times and combine all data
into one big dataset

https://www.basketball-reference.com/draft/NBA_2018.html

| | | | Round 1 | | | | | Totals | | | | | Shooting | | | Per Game | | | | Averages | |
|----|----|-----|-------------------------------|---------------------------------------|-----|----|------|--------|-----|-----|------|------|----------|------|------|----------|-----|-----|-----|----------|--|
| Rk | Pk | Tm | Player | College | Yrs | G | MP | PTS | TRB | AST | FG% | 3P% | FT% | MP | PTS | TRB | AST | WS | WS/ | | |
| 1 | 1 | PHO | Deandre Ayton | University of Arizona | 1 | 71 | 2183 | 1159 | 729 | 125 | .585 | .000 | .746 | 30.7 | 16.3 | 10.3 | 1.8 | 5.8 | .1 | | |
| 2 | 2 | SAC | Marvin Bagley | Duke University | 1 | 62 | 1567 | 923 | 471 | 62 | .504 | .313 | .691 | 25.3 | 14.9 | 7.6 | 1.0 | 3.6 | .1 | | |
| 3 | 3 | ATL | Luka Doncic | | 1 | 72 | 2318 | 1526 | 563 | 429 | .427 | .327 | .713 | 32.2 | 21.2 | 7.8 | 6.0 | 4.9 | .1 | | |

```
><td class="left " data-append-csv="aytonde01" data-stat="
"player" csk="Ayton,Deandre">...</td>
><td class="left " data-stat="college_name" csk="University of
Arizona">...</td>
<td class="right " data-stat="seasons">1</td>
<td class="right " data-stat="g">71</td>
<td class="right " data-stat="mp">2183</td>
<td class="right " data-stat="pts">1159</td>
<td class="right " data-stat="trb">729</td>
<td class="right " data-stat="ast">125</td>
<td class="right " data-stat="fg_pct">.585</td> == $0
<td class="right " data-stat="fg3_pct">.000</td>
<td class="right " data-stat="ft_pct">.746</td>
```

Draft Years

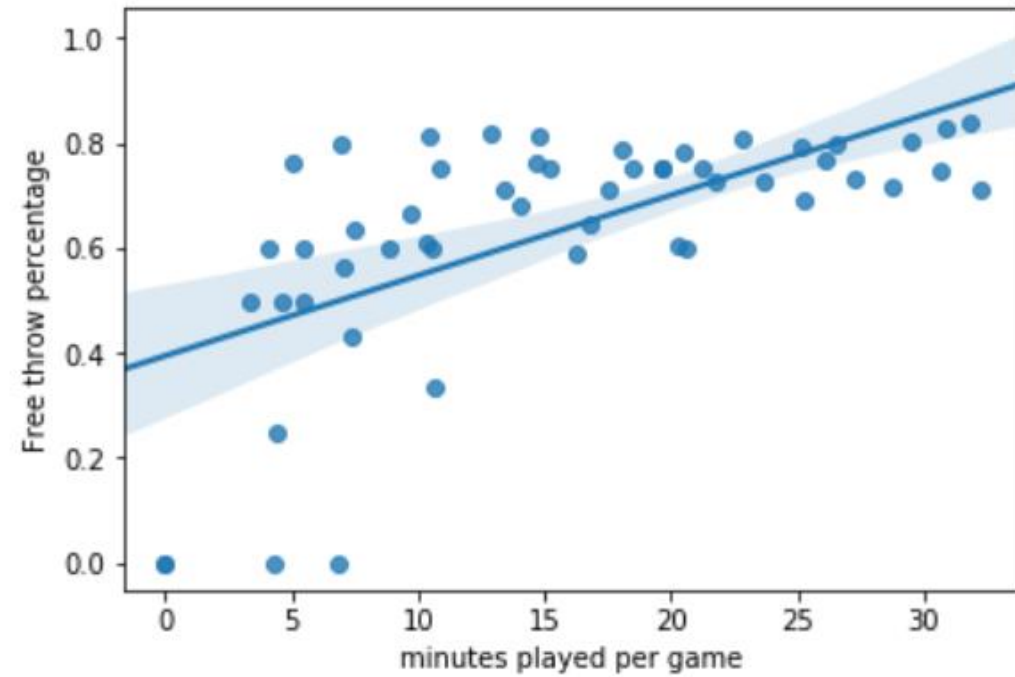
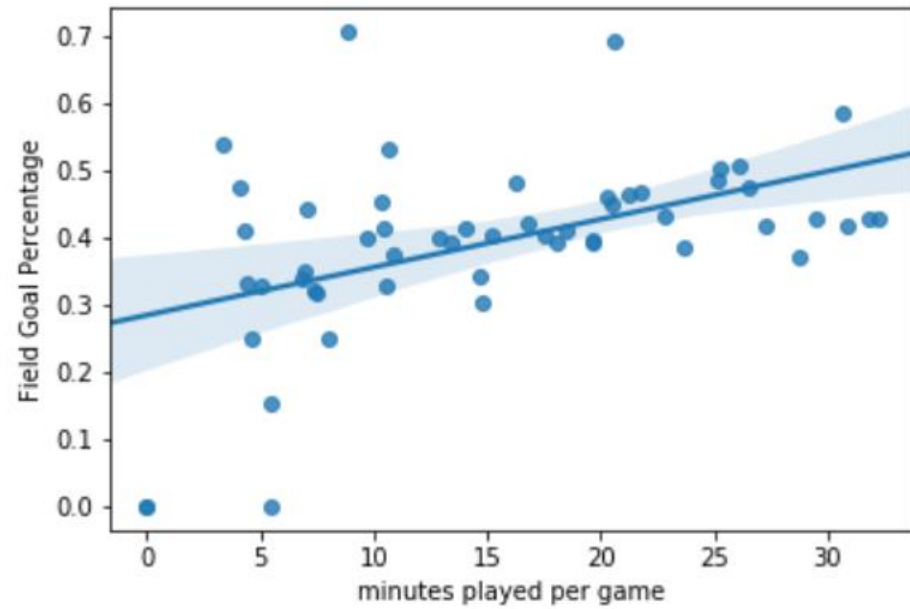
[Draft History](#)

[1947](#) [1948](#) [1949](#) [1950](#) [1951](#) [1952](#) [1953](#) [1954](#) [1955](#) [1956](#) [1957](#) [1958](#) [1959](#) [1960](#) [1961](#) [1962](#) [1963](#) [1964](#) [1965](#) [1966](#) [1967](#) [1968](#)
[1969](#) [1970](#) [1971](#) [1972](#) [1973](#) [1974](#) [1975](#) [1976](#) [1977](#) [1978](#) [1979](#) [1980](#) [1981](#) [1982](#) [1983](#) [1984](#) [1985](#) [1986](#) [1987](#) [1988](#) [1989](#) [1990](#)
[1991](#) [1992](#) [1993](#) [1994](#) [1995](#) [1996](#) [1997](#) [1998](#) [1999](#) [2000](#) [2001](#) [2002](#) [2003](#) [2004](#) [2005](#) [2006](#) [2007](#) [2008](#) [2009](#) [2010](#) [2011](#) [2012](#)
[2013](#) [2014](#) [2015](#) [2016](#) [2017](#) [2018](#)

Structured Dataset

| Player | College | Yrs | G | MP | PTS | TRB | AST | ... | 3P% | FT% | MP | PTS | TRB | AST | WS | WS/48 | BPM | VORP |
|---------------|-----------------------|-----|----|------|------|-----|-----|-----|------|------|------|------|------|-----|-----|-------|------|------|
| Deandre Ayton | University of Arizona | 1 | 71 | 2183 | 1159 | 729 | 125 | ... | .000 | .746 | 30.7 | 16.3 | 10.3 | 1.8 | 5.8 | .128 | 0.2 | 1.2 |
| Marvin Bagley | Duke University | 1 | 62 | 1567 | 923 | 471 | 62 | ... | .313 | .691 | 25.3 | 14.9 | 7.6 | 1.0 | 3.6 | .110 | -1.8 | 0.1 |
| Luka Doncic | | 1 | 72 | 2318 | 1526 | 563 | 429 | ... | .327 | .713 | 32.2 | 21.2 | 7.8 | 6.0 | 4.9 | .101 | 4.1 | 3.6 |

Data Exploratory



Proposal

- Use multiple linear regression build a prediction model
- Include players information to extend the analysis
- Based on last 10 years data to build a simulator to predict 2019 NBA draft