NBA Draft (College) Basketball Prediction

418 Final Project

Yi Chen

Web Scraping with python to extract 49 years data from 1969 -2017 via basketball reference. com website

Mean Imputation: Cleaning the data and get rid of missing values

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

			Round 1			Totals					Shooting			Per Game					A
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

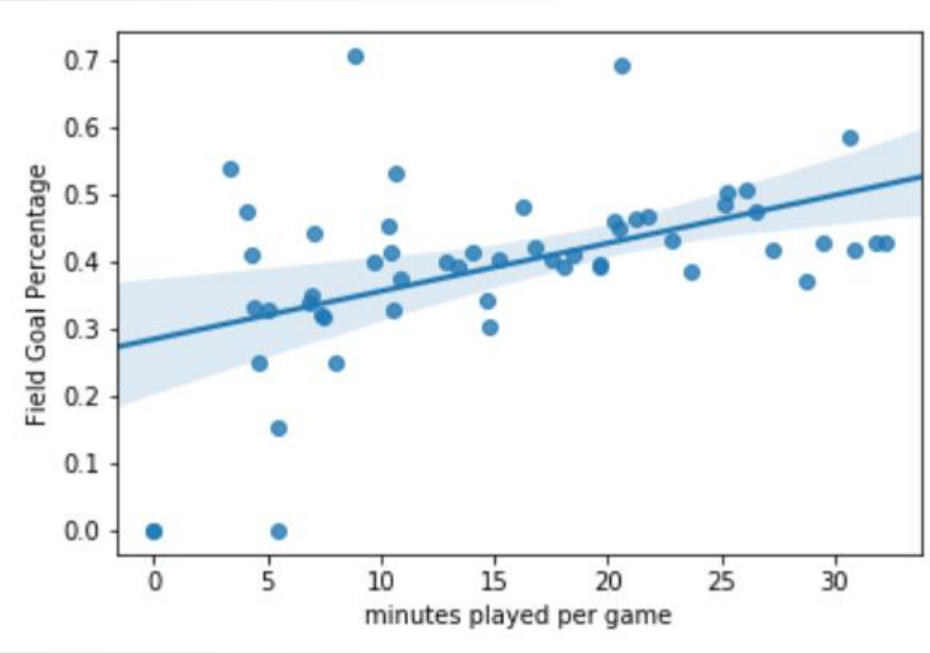
Draft History

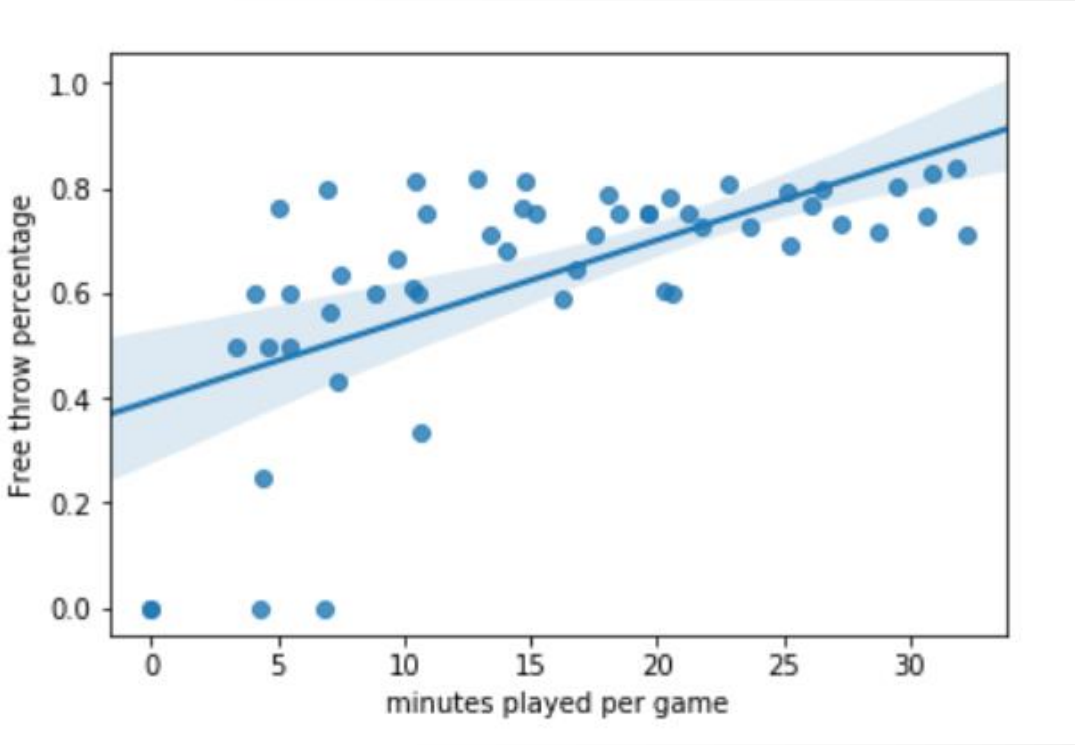
1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968
1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990
1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 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 Analysis



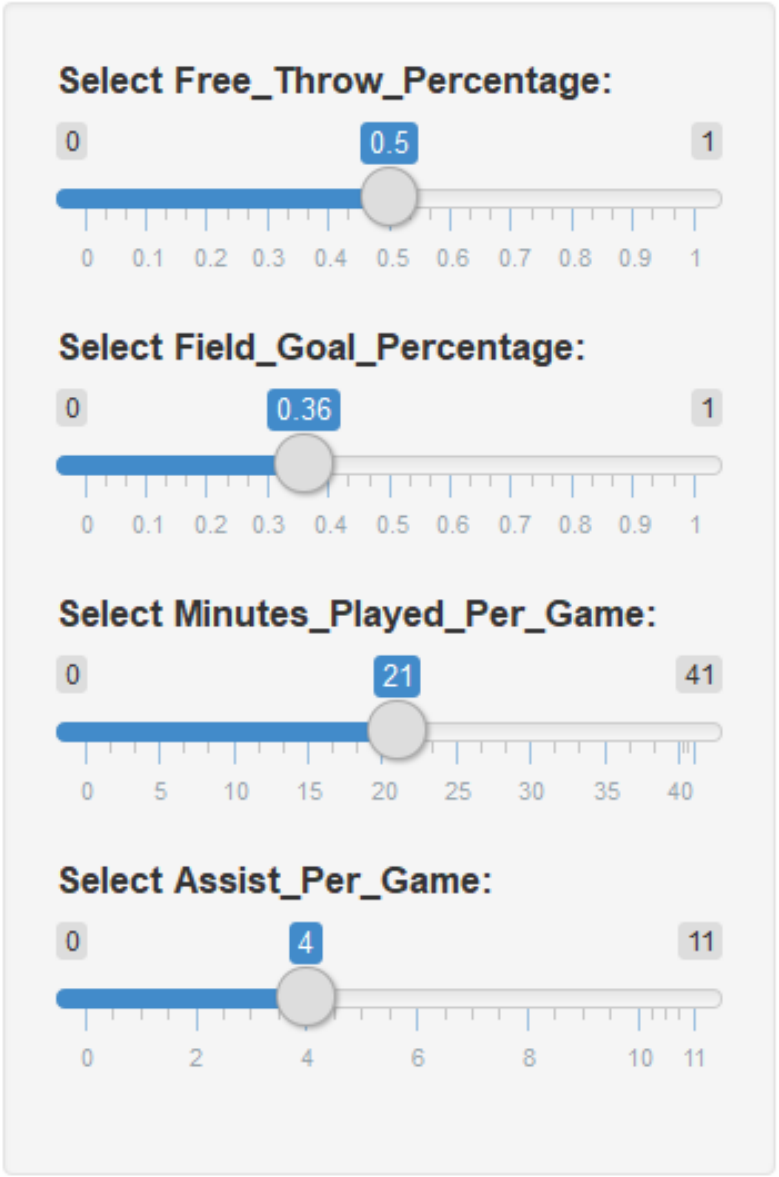


Use Multiple linear regression model to predict points per game

```
Call:
lm(formula = Points_Per_Game ~ Free_Throw_Percentage + Field_Goal_Percentage +
    Minutes_Played_Per_Game + Assist_Per_Game, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4402 -1.1606 -0.0563  0.9931 12.2634

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            -4.559289   0.313300 -14.552  < 2e-16 ***
Free_Throw_Percentage   2.761866   0.336593   8.205 3.58e-16 ***
Field_Goal_Percentage   2.491343   0.499687   4.986 6.58e-07 ***
Minutes_Played_Per_Game 0.480739   0.006443  74.609  < 2e-16 ***
Assist_Per_Game         0.082877   0.035846   2.312   0.0209 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.926 on 2584 degrees of freedom
Multiple R-squared:  0.8473,    Adjusted R-squared:  0.8471
F-statistic:  3585 on 4 and 2584 DF,  p-value: < 2.2e-16
```

Deploy App in Shiny

Predicting pointspergame



8.15