How Do New Advanced Player Tracking Stats Affect and Predict Team Performance and Efficiency?

Team Oklahoma City Thunder: Evan DeCastros, Junyi Huang, Yichen Pan, Sanjit Sisodiya

Overview

At the start of the 2013-14 season, the NBA implemented player tracking systems into team arenas

This has provided a wealth of new, smarter data which can give more insight into the different styles that teams play than traditional box scores and even advanced statistics give

Stats such as passes, secondary assists, drives, dribbles, seconds per touch, distance traveled, etc. can show which teams play a more unselfish style of play with ball and player movement compared to other teams who have ball dominators and are more isolation based

We have pulled these stats from nba.com, have adjusted per 100 possessions, and will explore their correlations with team performance

Hypotheses

There is a positive correlation between unselfish play and team performance

(higher assist numbers, more ball and player movement)

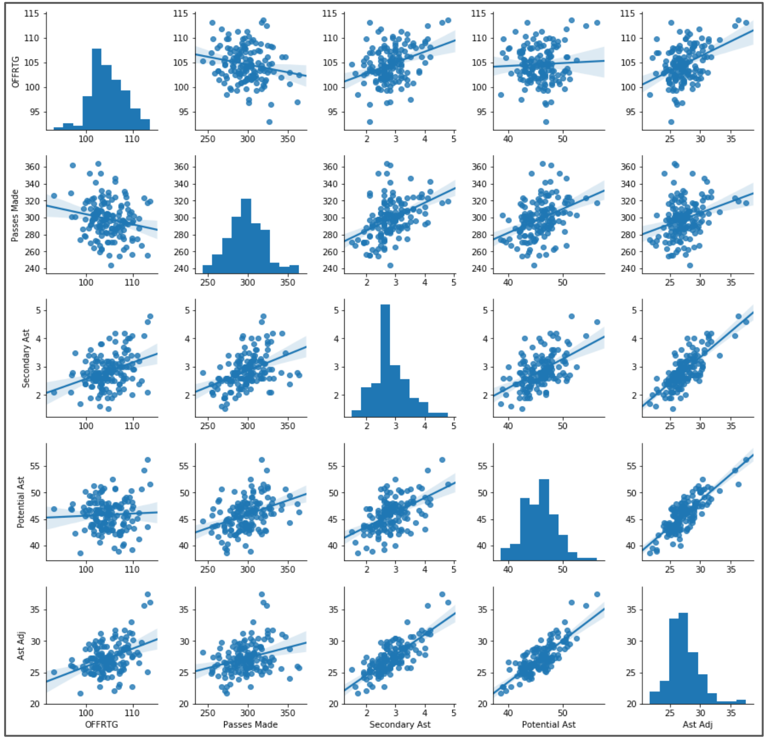
There is a negative correlation between isolation-based play and team performance

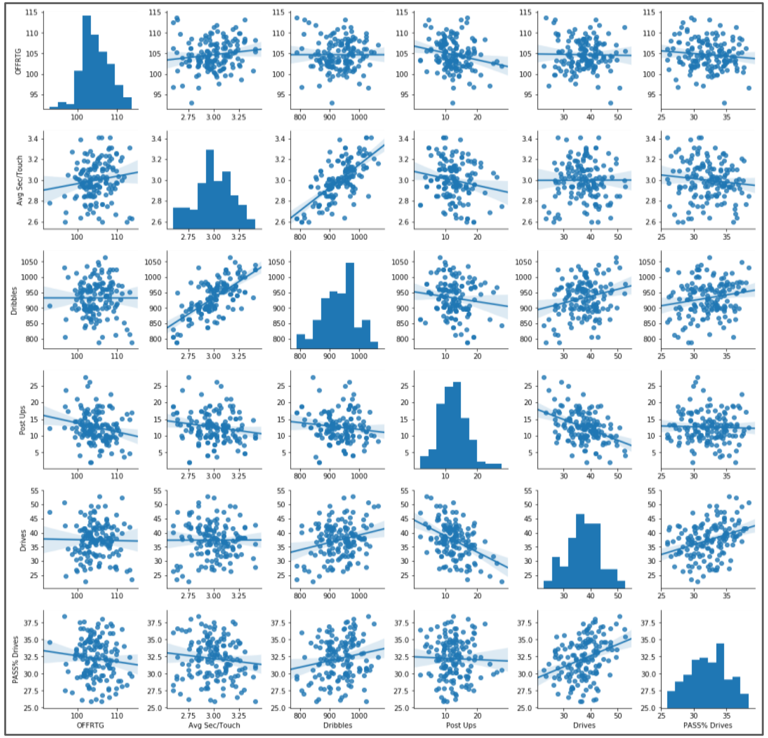
(post ups, more dribbling and driving)

Correlation between offensive rating and win percentage is .77 – regression will target offensive rating since all independent variables are offensive-based

Pair Plots of Relationships with OFFRTG

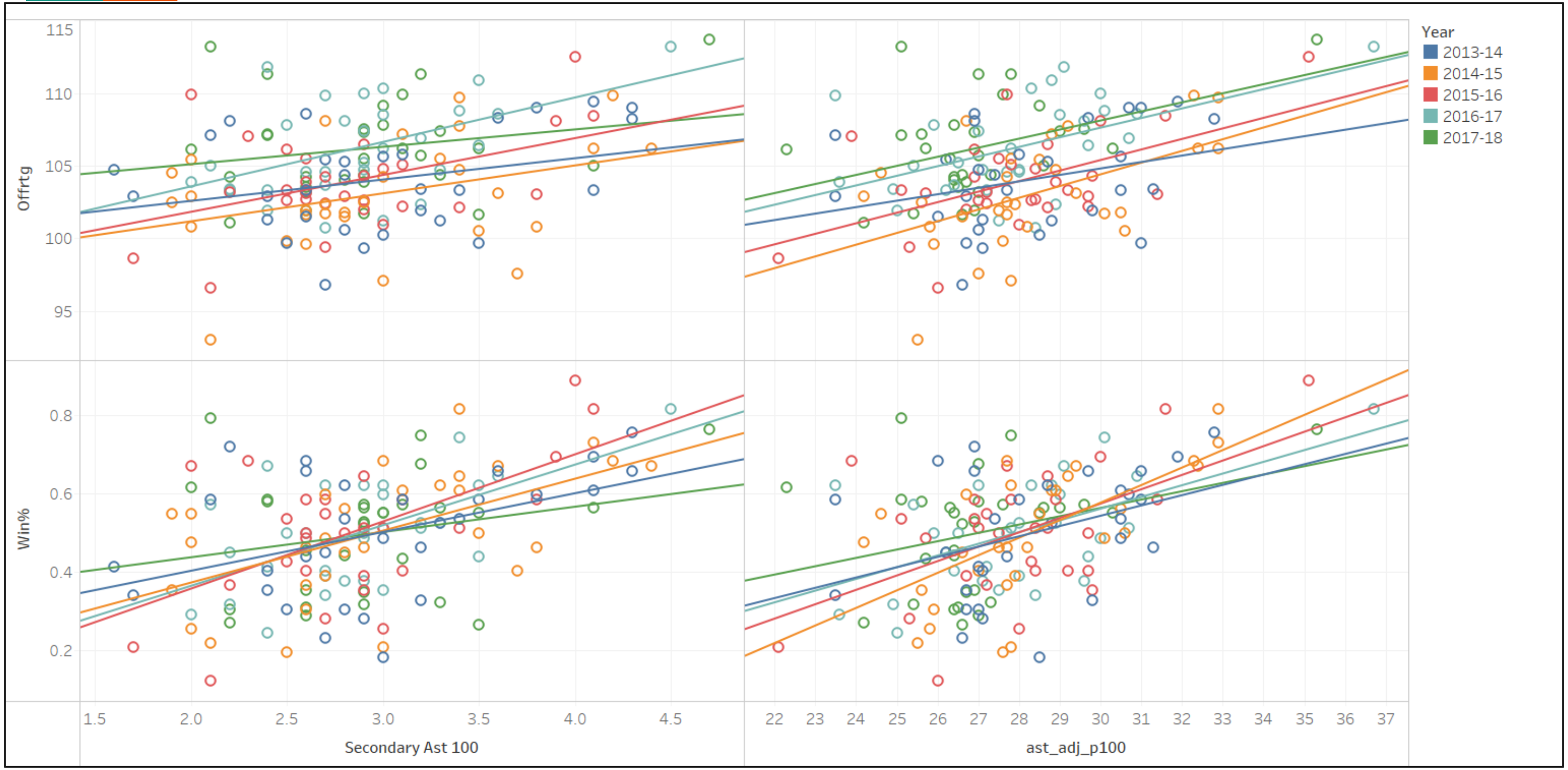
Offensive Rating vs Passing Variables



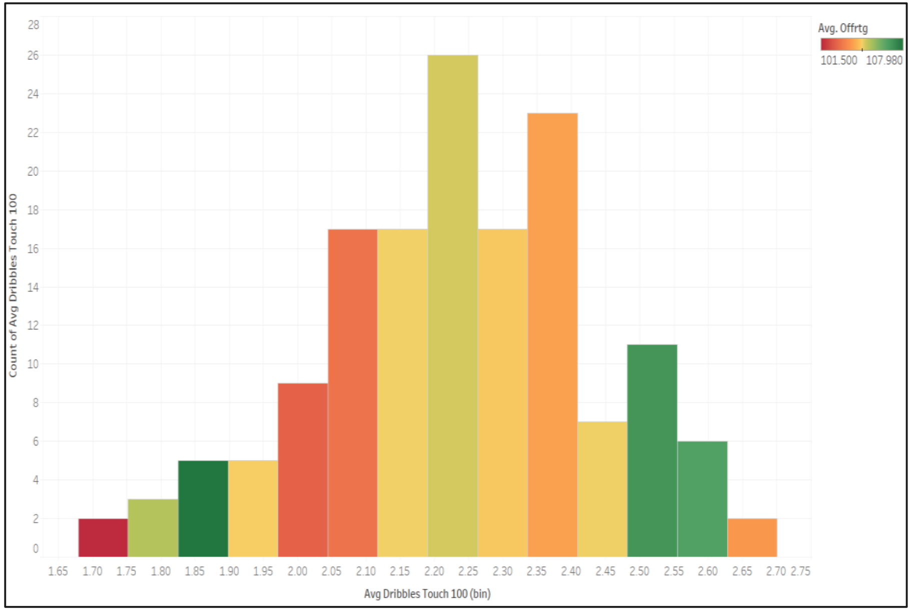


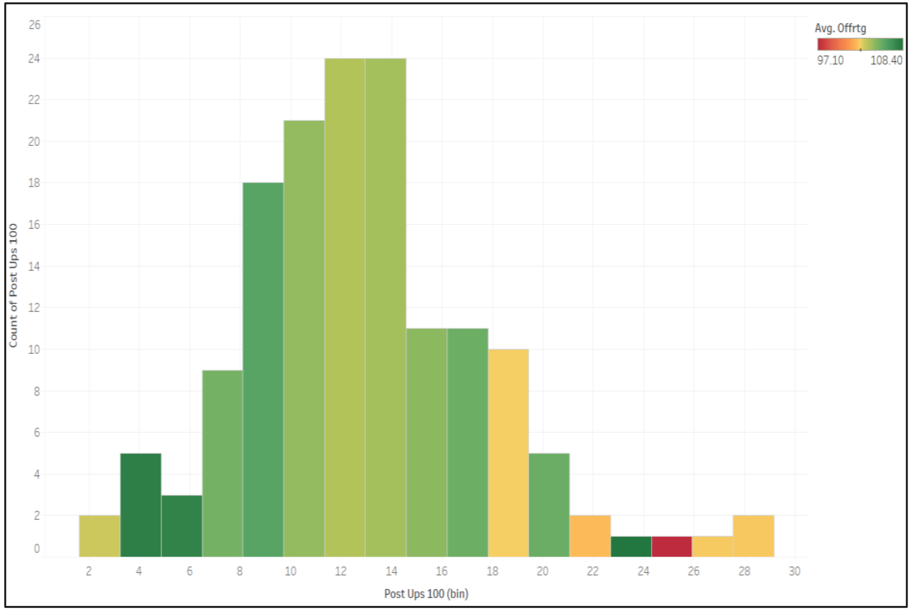
Offensive Rating vs Isolation Variables

Secondary and Adjusted Assists vs Win % and OFFRTG

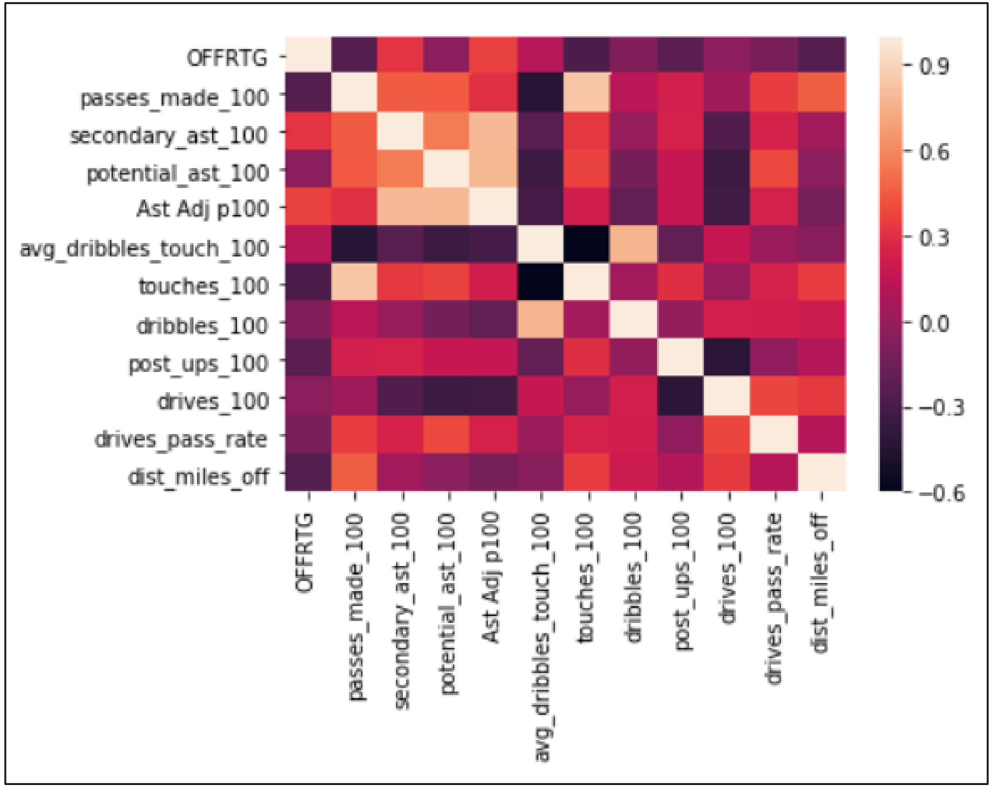


Average OFFRTG by Dribbles per Touch and Post Ups





Correlation Matrix of Variables



Variable Selection and Descriptive Statistics

	OFFRTG (dependent)	passes_made_100	secondary_ast_100	potential_ast_100	ast_adj_100	touches_100
count	150	150	150	150	150	150
mean	104.669	305.387	2.908	46.877	27.904	429.149
std	3.604	25.059	0.612	2.979	2.397	25.848
min	93	247.4	1.6	39.5	22.1	374.7
25%	102.3	289.15	2.6	44.8	26.525	412.125
50%	104.4	303.3	2.9	46.95	27.7	427.65
75%	107.2	320.25	3.2	49	29.175	445.3
max	113.7	392.4	4.7	55.1	36.7	517
	avg_dribbles_touch_100	**dribbles_100**	**post_ups_100**	**drives_100**	**drives_pass_rate**	**dist_miles_off**
count	150	150	150	150	150	150
mean	2.232	954.742	12.885	38.257	32.204	9.088
std	0.206	70.826	4.711	6.087	2.949	0.299
min	1.71	770.2	2.1	22.9	25.9	8.5
25%	2.09	908.625	9.8	34.225	30	8.89
50%	2.25	955.15	12.55	38.55	32.25	9.065
75%	2.3675	1005.35	15.175	41.975	34.5	9.2975
max	2.69	1117.3	28.9	54.1	38.4	9.89

Methodology

Linear Regression and Clustering

Identify statistically significant tracking variables in predicting offensive rating with linear regression in R

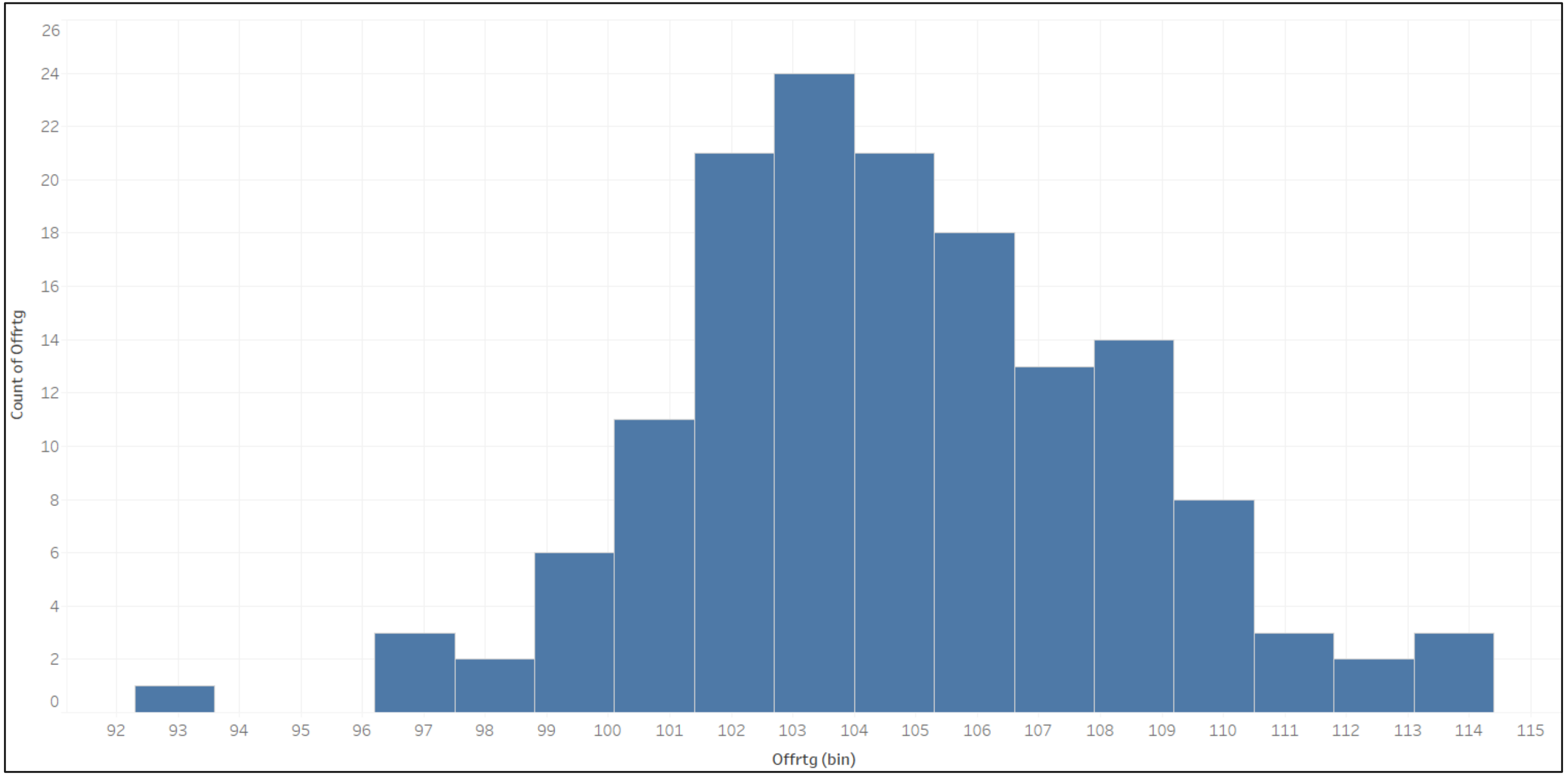
- R-squared and predictive capability may be limited without including four factors/efficiency

Cluster teams based on their style of play using variables highlighted from regression

- Use two-step clustering node in SPSS Modeler to determine optimal number of clusters

Analyze team performance trends within clusters

Normal Distribution of Target Variable OFFRTG



Multi-Collinearity of Independent Variables

Variables	VIF
passes_made_100	5.098441
secondary_ast_100	3.878229
potential_ast_100	4.621028
Ast.Adj.p100	6.368055
touches_100	105.462890
avg_dribbles_touch_100	257.436510
dribbles_100	166.069136
post_ups_100	1.514354
drives_100	2.828277
drives_pass_rate	2.231825
dist_miles_off	1.739713

Variables	VIF
passes_made_100	5.096941
secondary_ast_100	3.814870
potential_ast_100	4.588845
Ast.Adj.p100	6.354035
touches_100	3.851742
dribbles_100	1.241971
post_ups_100	1.512656
drives_100	2.819154
drives_pass_rate	2.222295
dist_miles_off	1.738365

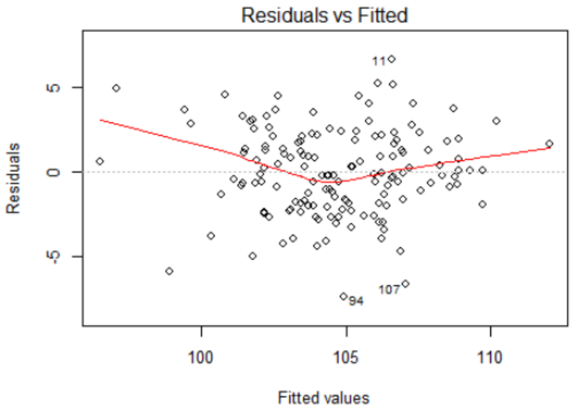
Variables	VIF
passes_made_100	2.624652
secondary_ast_100	3.767910
potential_ast_100	4.497801
Ast.Adj.p100	6.097236
avg_dribbles_touch_100	1.468167
post_ups_100	1.427440
drives_100	2.732341
drives_pass_rate	2.200982
dist_miles_off	1.713535

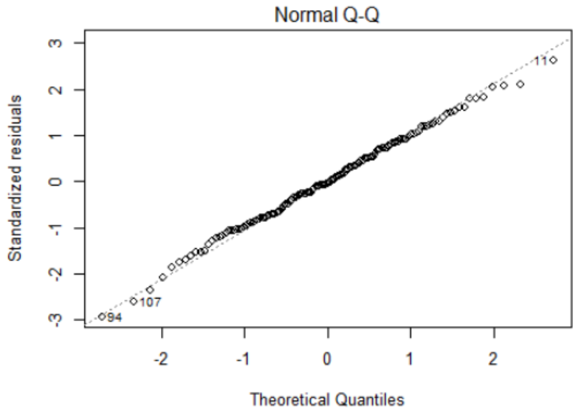
- Extremely high multi-collinearity between dribbles per touch and dribbles, touches
  - Use in separate models
- Also some multi-collinearity between adjusted assists and other passing variables – it is the summation of actual assists, secondary assists and free throw assists

Linear Regression

```
Call:
lm(formula = OFFRTG ~ passes_made_100 + secondary_ast_100 + potential_ast_100 +
    Ast.Adj.p100 + avg_dribbles_touch_100 + post_ups_100 + drives_100 +
    drives_pass_rate + dist_miles_off, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-7.3221 -1.8156 -0.1135  1.7843  6.5858

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            119.07885    9.06781  13.132  < 2e-16 ***
passes_made_100         -0.01854    0.01371  -1.352 0.178415
secondary_ast_100        0.83941    0.67224   1.249 0.213865
potential_ast_100       -0.81799    0.15095  -5.419 2.55e-07 ***
Ast.Adj.p100             1.32421    0.21846   6.062 1.18e-08 ***
avg_dribbles_touch_100   1.63482    1.24537   1.313 0.191426
post_ups_100            -0.18150    0.05378  -3.375 0.000955 ***
drives_100              -0.01129    0.05758  -0.196 0.844874
drives_pass_rate        -0.04008    0.10665  -0.376 0.707661
dist_miles_off          -1.03249    0.92955  -1.111 0.268580
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.588 on 140 degrees of freedom
Multiple R-squared:  0.5156,    Adjusted R-squared:  0.4844
F-statistic: 16.56 on 9 and 140 DF,  p-value: < 2.2e-16
```

R-Squared: .4844





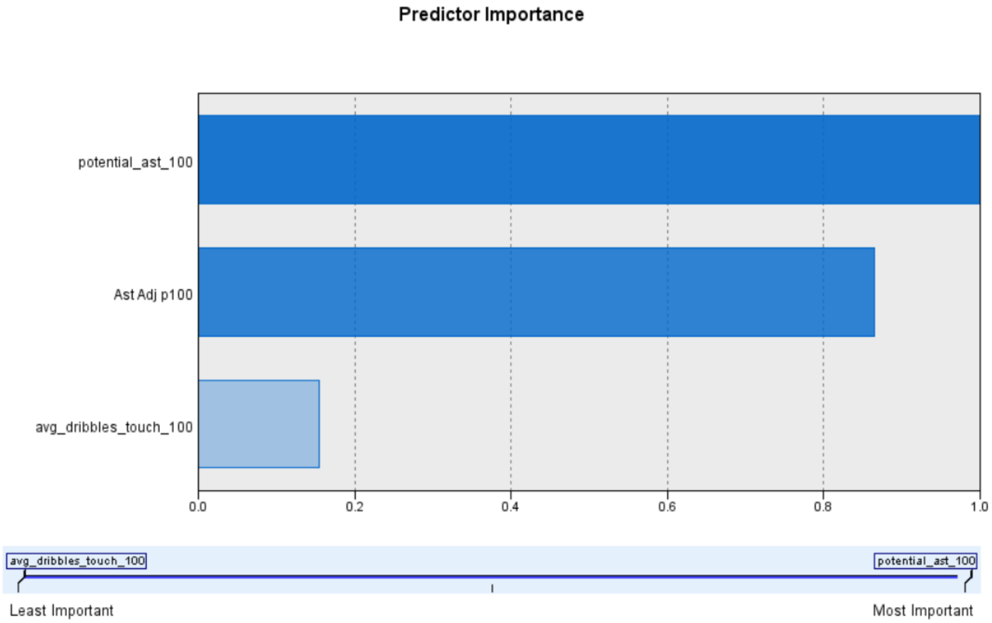
Model explains over 48% of variance in OFFRTG

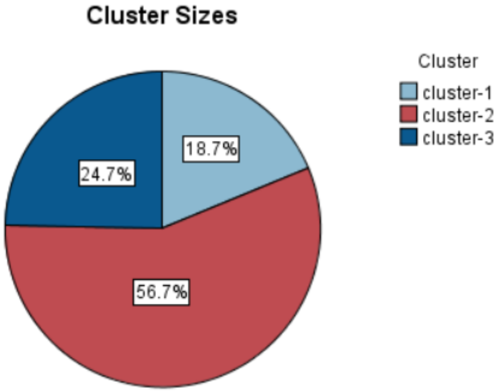
Clustering

- Adjusted Assists and Potential Assists were two most significant variables
  - Also a good reflection of team ball movement
- Post ups were statistically significant but did not fit neatly into either category of play, and had lowest importance in cluster creation
  - Post ups had a negative coefficient in the linear regression
  - Clusters showed ball movement teams having higher number of post ups per 100
- Substituted avg dribbles/touch which is a better reflection of isolation teams' style of play

	Cluster-1	Cluster-2	Cluster-3
Cluster Description	Ball Movement	Moderate	Isolation
Potential Ast Mean	50.77	47.23	43.12
Adjusted Ast Mean	31.43	27.82	25.43
Avg Dribbles/Touch Mean	2.09	2.22	2.36

Clustering Model Details

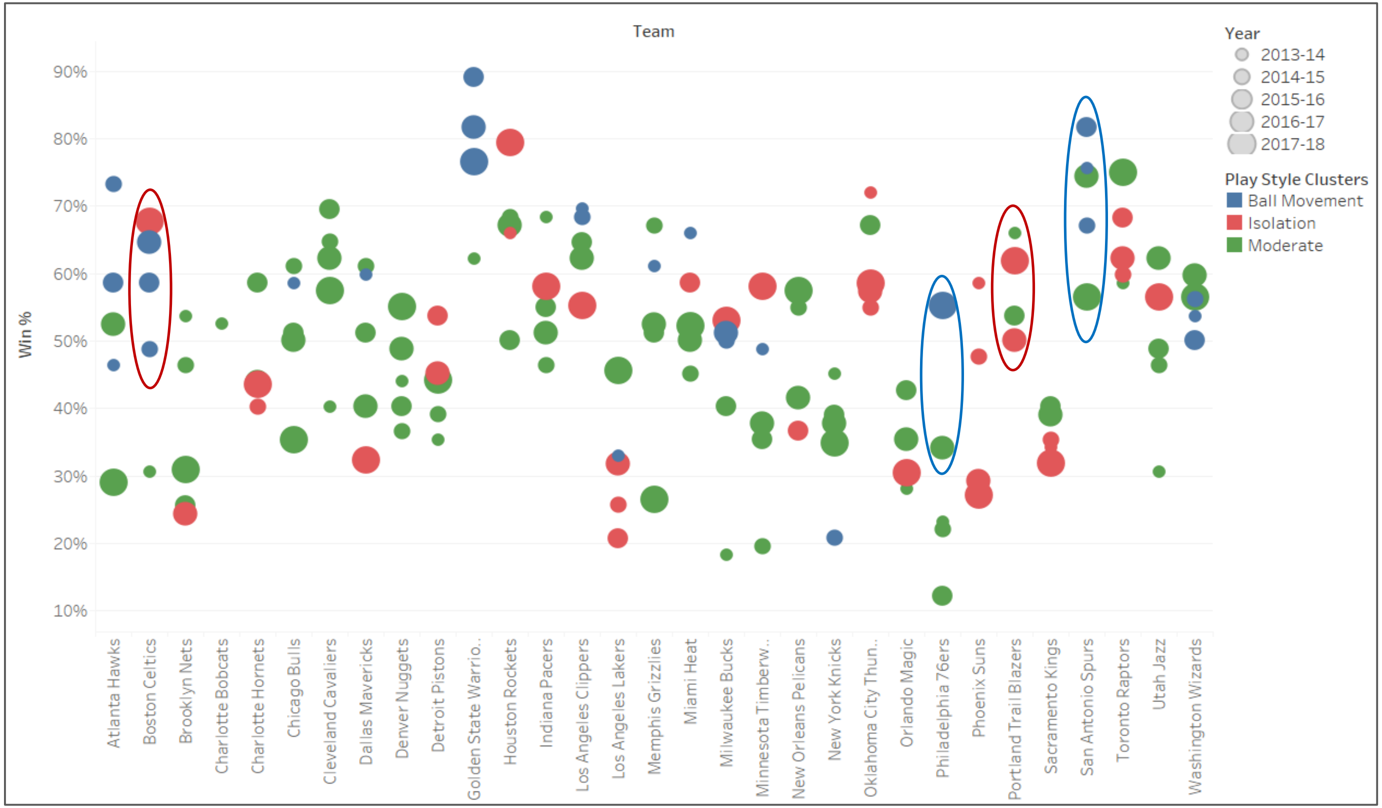




Model Summary

Algorithm	TwoStep
Inputs	3
Clusters	3

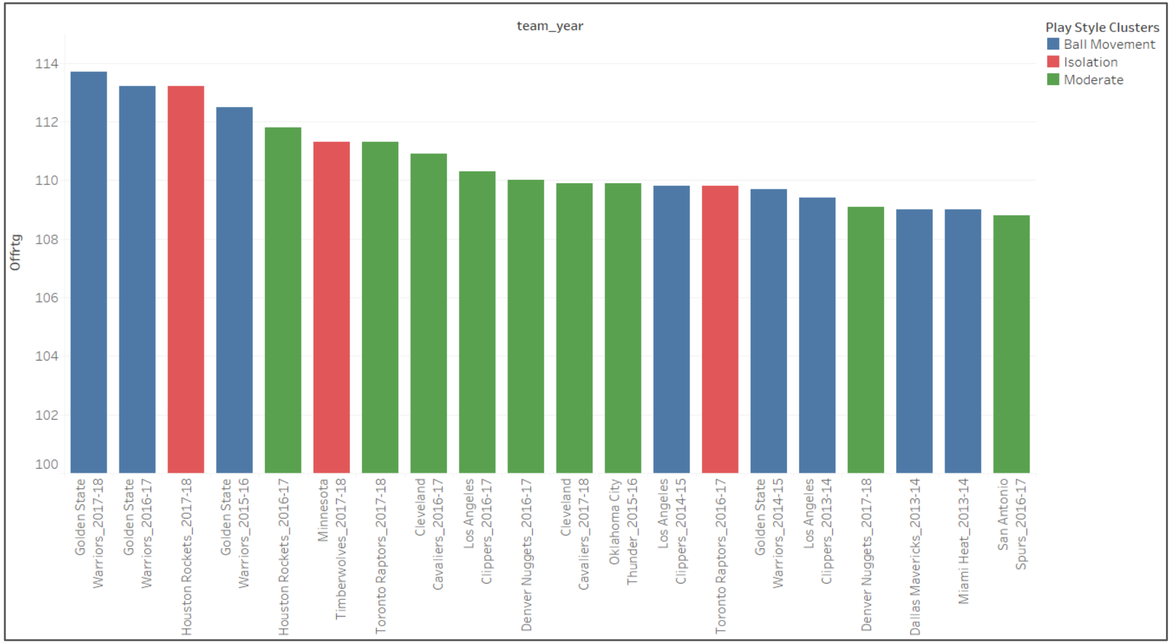
Distribution of Clusters Among Teams Since 2013-14

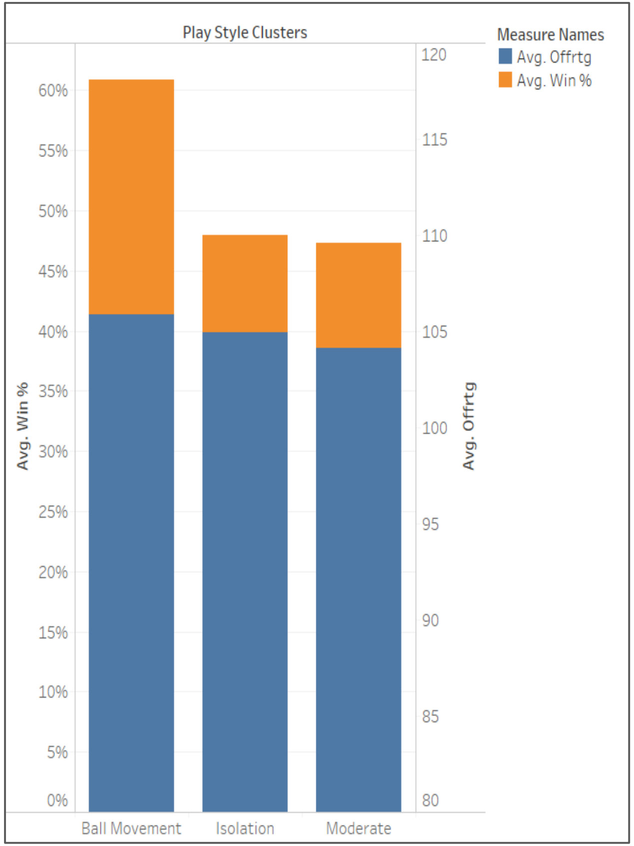


Matches eye test anecdotally:

- Many teams played the same style for majority of years (Thunder iso based, Warriors ball movement based)
- Spurs with Duncan were ball movement based (2013-15)
- Celtics and Blazers have become more iso based with acquisition of Kyrie Irving and emergence of Lillard, McCollum respectively – guards known to dominate the ball
- 76ers have become a ball movement team with addition of stud rookie Ben Simmons

Top 20 Offensive Teams and Performance Within Clusters





- Top 20 offensive teams of last 5 years consist of 8 ball movement teams and only 3 iso teams
- Win % is more affected by style than OFFRTG
- % of teams with records over .500: BM = 75%, Iso = 51%, Moderate = 45%

ANOVA Comparing Performance of Clusters

```
> offrtg.aov<- aov(clusters$OFFRTG ~ clusters$play_style_clusters)
> summary(offrtg.aov)
                              Df Sum Sq Mean Sq F value Pr(>F)
clusters$play_style_clusters   2   68.2   34.11   2.685 0.0716 .
Residuals                    147 1867.6   12.70
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(offrtg.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = clusters$OFFRTG ~ clusters$play_style_clusters)

$`clusters$play_style_clusters`
                             diff       lwr        upr     p adj
Isolation-Ball Movement -0.9531853 -3.067086 1.16071514 0.5356586
Moderate-Ball Movement  -1.7505462 -3.589448 0.08835533 0.0657212
Moderate-Isolation      -0.7973609 -2.459538 0.86481633 0.4936515
```

```
> wins.aov<- aov(clusters$win_pct ~ clusters$play_style_clusters)
> summary(wins.aov)
                              Df Sum Sq Mean Sq F value   Pr(>F)
clusters$play_style_clusters   2 0.4088 0.20438   9.616 0.000119 ***
Residuals                    147 3.1245 0.02126
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(wins.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = clusters$win_pct ~ clusters$play_style_clusters)

$`clusters$play_style_clusters`
                                diff         lwr         upr     p adj
Isolation-Ball Movement -0.129012373 -0.21547605 -0.04254870 0.0015856
Moderate-Ball Movement  -0.135850906 -0.21106645 -0.06063536 0.0001005
Moderate-Isolation      -0.006838533 -0.07482563  0.06114856 0.9692235
```

- ANOVA shows that there is not a statistically significant difference for offensive rating between clusters, but there is for win %
- Previous chart showed 61% average win % for ball movement teams, and only 48% and 47% for isolation and moderate teams

  - Ball movement average win % is still 57% when excluding Golden State Warriors

Conclusions, Implications and Future Research

The most significant tracking variables in predicting offensive rating were adjusted assists and potential assists

Teams who play with more ball movement win a higher percentage of games on average

Not always the case as shown by the Houston Rockets with James Harden and Chris Paul

Players who are efficient in scoring out of isolation style plays still have a positive impact on team performance

The NBA is still a player driven league, and identifying the effect of specific players/types of players in this regard is an important next step

Data Source

https://stats.nba.com/teams/passing/
https://stats.nba.com/teams/drives/
https://stats.nba.com/teams/touches/
https://stats.nba.com/teams/speed-distance/

- For 2017-18 season, data is as of 3/16/2018

https://stats.nba.com/help/glossary/

- Glossary including full descriptions of variables