Analysis and Prediction of NBA MVP Award Winners

Mateen Hussaini

On-Uma Lomsomboot

Goals

- Understand playstyle of MVP Candidates vs. Non-Candidates with **Principal Component Analysis (PCA)**
- Identify which basic stats are strongly connected to NBA player receiving MVP votes with **Decision Trees**
- Determine minimum qualifications to be considered MVP and MVP candidates **with Rule-Based Learning**
- Create an algorithm to predict 2022 MVP and MVP candidates using 2022 stats with **Logistic Regression**
- Compare using past 10, 15, 31 years of data affects prediction of MVP
- Research and develop algorithm that can predict 2022 MVP using 2021 stats (Failed)
  - Didn't have time to research and develop algorithm

Programs Used

- Google Colab (Coding on web browser)
- R (Principal Component Analysis)
- Excel (tables and graph)
- Python
    - RuleFit (Rule-Based Learning)
    - Scikit Learn (Decision Trees and Variable Importance)
    - NumPy (Arrays)
    - Matplotlib (Graphs and Plots)

Data

- 14092 observations from the 1991 season to 2021 season
  - 31 MVPS, 440 MVP Candidates, 13621 non-candidates
- 44 variables in our data (categorical and quantatative)
  - Included player, team, and voting stats
- Scraped from Basketball-Reference by Kaggle user Vivo Vinco
- Trimmed down 44 variables to 15 variables(all quantitative)
  - Removed unrelated variables e.g Pts Won
  - Removed double counted variables e.g ORB + DRB = TRB, so we only kept TRB
- Response Variable is called MVP
  - Classified players into 3 groups: MVP, MVP Candidate, Non-Candidate

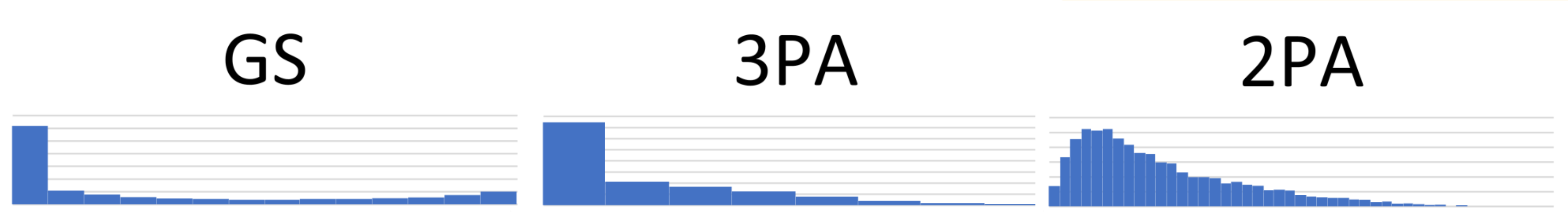
Variables

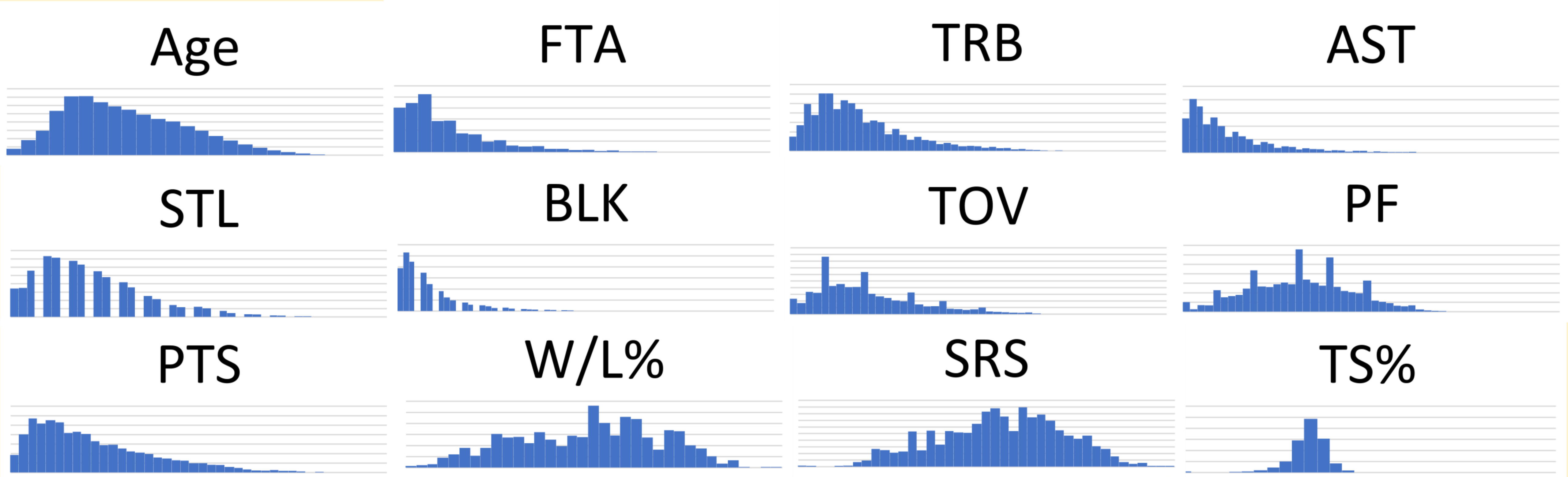
- Excluded MP
- Used TS% instead of individually using 3PA%, 2P%, and FT%
- Believed PTS, TS%, TRB, AST, W/L% would be most important variables
- W/L% and SRS are team stats

Age	Age of Player
GS	Games Started
3PA	3 Pointers Attempted 1 make = 3 points
2PA	2 Pointers Attempted 1 make = 2 points
FTA	Free Throws Attempted 1 make = 1 point
TRB	Total Rebounds
AST	Assist
STL	Steals
BLK	Blocks
TOV	Turnovers
PF	Personal Fouls
PTS	Points Scored
W/L%	Percentage of games teams has won
SRS	Simple Rating System from Basketball Reference. Used to measure value of wins e.g A 20 point win over the best team increases SRS e.g A 20 point lost to the worst team decreases SRS
TS%	Combination of % of FGA and FTA Describe how often player makes points attempt $\frac{PTS}{(2*FGA) + (0.44*FTA)}$

Variable Distribution 1991-2021 Stats

- Most of the data is skewed right





Data Pre-Processing

	Unnamed: 0	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	Pts Max	Share	Team	W	L	W/L%	GB	PS/G	PA/G	SRS
0	0	?an Tabak	C	24	HOU	37	0	4.9	0.6	1.4	...	0.0	0.0	Houston Rockets	47	35	0.573	15.0	103.5	101.4	2.32
1	1	Adrian Caldwell	PF	28	HOU	7	0	4.3	0.1	0.6	...	0.0	0.0	Houston Rockets	47	35	0.573	15.0	103.5	101.4	2.32
2	2	Carl Herrera	PF	28	HOU	61	26	21.8	2.8	5.4	...	0.0	0.0	Houston Rockets	47	35	0.573	15.0	103.5	101.4	2.32
3	3	Charles Jones	PF	37	HOU	3	0	12.0	0.3	1.0	...	0.0	0.0	Houston Rockets	47	35	0.573	15.0	103.5	101.4	2.32
4	4	Chucky Brown	SF	26	HOU	41	14	19.9	2.6	4.2	...	0.0	0.0	Houston Rockets	47	35	0.573	15.0	103.5	101.4	2.32
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14087	14087	Spencer Hawes	PF	28	MIL	54	1	14.8	2.5	5.1	...	0.0	0.0	Milwaukee Bucks	42	40	0.512	9.0	103.6	103.8	-0.45
14088	14088	Steve Novak	PF	33	MIL	8	0	2.8	0.3	0.9	...	0.0	0.0	Milwaukee Bucks	42	40	0.512	9.0	103.6	103.8	-0.45
14089	14089	Terrence Jones	PF	25	MIL	54	12	23.5	4.3	9.1	...	0.0	0.0	Milwaukee Bucks	42	40	0.512	9.0	103.6	103.8	-0.45
14090	14090	Thon Maker	C	19	MIL	57	34	9.9	1.5	3.2	...	0.0	0.0	Milwaukee Bucks	42	40	0.512	9.0	103.6	103.8	-0.45
14091	14091	Tony Snell	SG	25	MIL	80	80	29.2	3.1	6.8	...	0.0	0.0	Milwaukee Bucks	42	40	0.512	9.0	103.6	103.8	-0.45

14092 rows × 42 columns

Player Stats

MVP

	Rank	Player	Age	Tm	First	Pts Won	Pts Max	Share	G	MP	...	TRB	AST	STL	BLK	FG%	3P%	FT%	WS	WS/48	Year
0	1	Michael Jordan	27	CHI	77	891	960	0.928	82	37.0	...	6.0	5.5	2.7	1.0	0.539	0.312	0.851	20.3	0.321	1991
1	2	Magic Johnson	31	LAL	10	497	960	0.518	79	37.1	...	7.0	12.5	1.3	0.2	0.477	0.320	0.906	15.4	0.251	1991
2	3	David Robinson	25	SAS	6	476	960	0.496	82	37.7	...	13.0	2.5	1.5	3.9	0.552	0.143	0.762	17.0	0.264	1991
3	4	Charles Barkley	27	PHI	2	222	960	0.231	67	37.3	...	10.1	4.2	1.6	0.5	0.570	0.284	0.722	13.4	0.258	1991
4	5	Karl Malone	27	UTA	0	142	960	0.148	82	40.3	...	11.8	3.3	1.1	1.0	0.527	0.286	0.770	15.5	0.225	1991
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
469	11	Russell Westbrook	32	WAS	0	5	1010	0.005	65	36.4	...	11.5	11.7	1.4	0.4	0.439	0.315	0.656	3.7	0.075	2021
470	12	Ben Simmons	24	PHI	0	3	1010	0.003	58	32.4	...	7.2	6.9	1.6	0.6	0.557	0.300	0.613	6.0	0.153	2021
471	13T	James Harden	31	TOT	0	1	1010	0.001	44	36.6	...	7.9	10.8	1.2	0.8	0.466	0.362	0.861	7.0	0.208	2021
472	13T	LeBron James	36	LAL	0	1	1010	0.001	45	33.4	...	7.7	7.8	1.1	0.6	0.513	0.365	0.698	5.6	0.179	2021
473	13T	Kawhi Leonard	29	LAC	0	1	1010	0.001	52	34.1	...	6.5	5.2	1.6	0.4	0.512	0.398	0.885	8.8	0.238	2021

474 rows × 21 columns

	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	...	W	L	W/L%	GB	PS/G	PA/G	SRS	TS%	Rank	MVP
0	?an Tabak	C	-0.886218	HOU	-0.716531	-0.912480	-1.501674	-1.133253	-1.180332	0.132625	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.021818	0	0
1	Adrian Caldwell	PF	0.174886	HOU	-1.903423	-0.912480	-1.558538	-1.353423	-1.351556	-1.999926	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	-1.687118	0	0
2	Carl Herrera	PF	0.174886	HOU	0.232983	-0.044511	0.100016	-0.164503	-0.324215	0.867987	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.293066	0	0
3	Charles Jones	PF	2.562368	HOU	-2.061675	-0.912480	-0.828774	-1.265355	-1.265944	-1.127996	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	-1.371853	0	0
4	Chucky Brown	SF	-0.355666	HOU	-0.558279	-0.445112	-0.080055	-0.252571	-0.581050	1.708401	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	1.134430	0	0
5	Clyde Drexler	SG	1.235989	HOU	0.826429	1.591276	1.436338	1.905099	2.008705	0.216666	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.629882	14.0	2
6	Hakeem Olajuwon	C	1.235989	HOU	0.668176	1.491126	1.787004	3.490327	3.121658	0.804956	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.449002	5.0	2
7	Kenny Smith	PG	0.440161	HOU	1.024244	1.791576	0.412772	0.143736	0.082441	0.458285	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	1.288071	0	0
8	Mario Elie	SF	0.970713	HOU	1.024244	-0.478496	0.251656	-0.076434	-0.195797	0.615863	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	1.215131	0	0
9	Pete Chilcutt	PF	-0.355666	HOU	0.509924	-0.344962	-0.089533	-0.472741	-0.452632	0.048583	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.069118	0	0
10	Robert Horry	PF	-0.886218	HOU	0.351672	1.123908	1.104627	0.275838	0.317873	0.069594	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.391240	0	0
11	Sam Cassell	PG	-0.620942	HOU	1.063807	-0.879096	0.213746	-0.032400	0.061038	-0.140510	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	0.389371	0	0
12	Tim Breaux	SF	-0.886218	HOU	-0.518716	-0.845713	-1.198395	-0.913082	-0.859288	-0.718294	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	-0.806760	0	0
13	Tracy Murray	SF	-1.151494	HOU	-0.043959	-0.812330	-1.056233	-0.604844	-0.559647	-0.340108	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	-0.006432	0	0
14	Vernon Maxwell	SG	0.440161	HOU	0.351672	0.890224	1.047762	0.716179	1.109782	-0.487181	...	0.473361	-0.473361	0.472185	-0.112700	0.431817	-0.006957	0.509455	-0.095708	0	0
15	A.C. Green	SF	0.970713	PHO	1.063807	0.823457	1.142537	0.275838	0.125247	0.668388	...	1.391371	-1.391371	1.394055	-1.231418	1.862663	1.172556	0.837225	0.843114	0	0
16	Aaron Swinson	SF	-0.886218	PHO	-1.824297	-0.912480	-1.425854	-0.913082	-1.051915	1.214657	...	1.391371	-1.391371	1.394055	-1.231418	1.862663	1.172556	0.837225	0.814825	0	0
17	Antonio Lang	SF	-1.416770	PHO	-1.705608	-0.912480	-1.549061	-1.265355	-1.308750	-0.424150	...	1.391371	-1.391371	1.394055	-1.231418	1.862663	1.172556	0.837225	-0.344349	0	0
18	Charles Barkley	PF	0.970713	PHO	0.509924	1.290825	1.351041	2.169304	2.115720	0.479295	...	1.391371	-1.391371	1.394055	-1.231418	1.862663	1.172556	0.837225	0.568083	6.0	2

All players with MVP Column

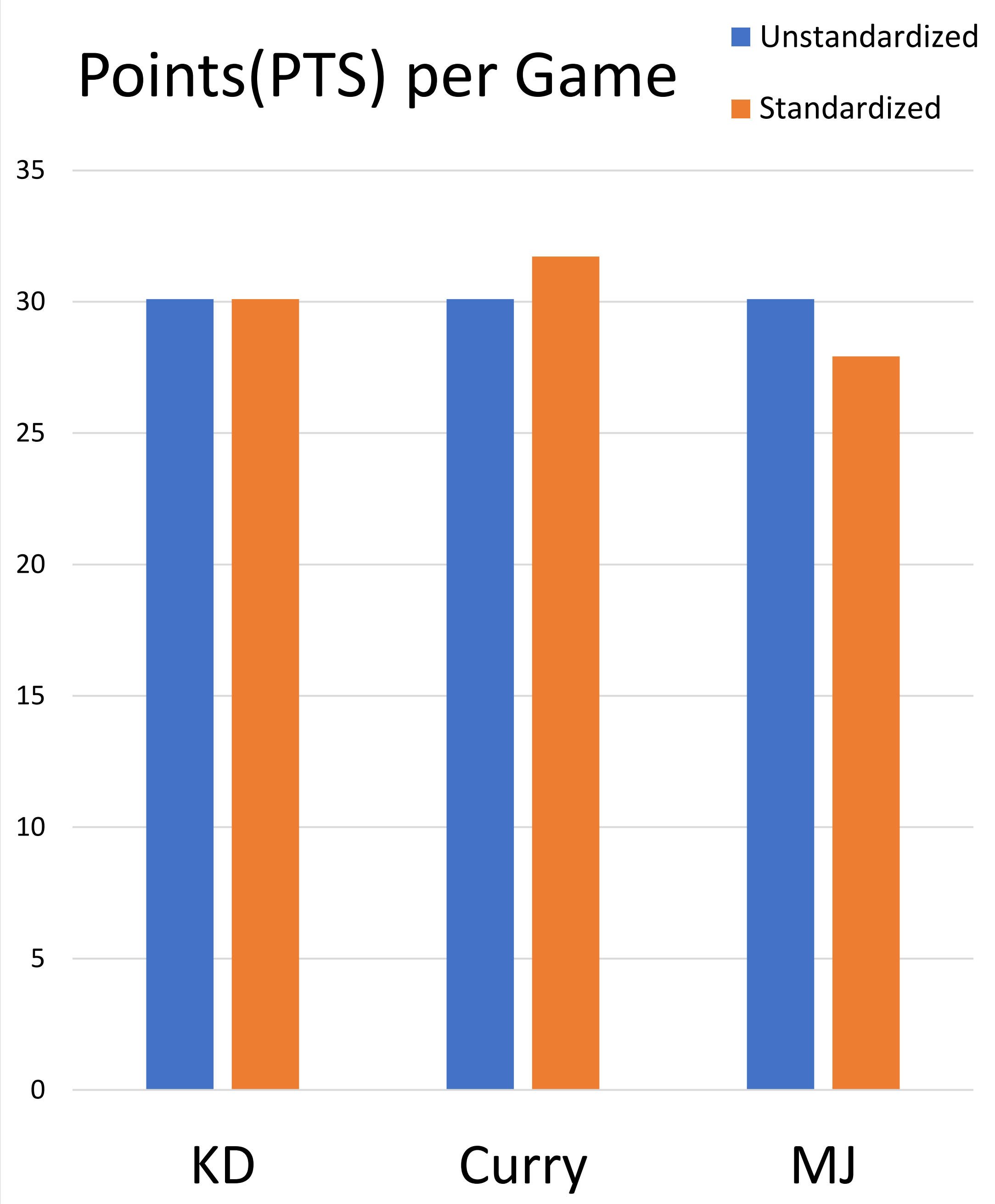
0 = Non-Candidate

1 = MVP

2 = MVP Candidate

Standardizing Data

- Rule changes can affect the magnitude of stats season by season
- **We separated data by season into separate data frames**
- **Standardized each data frame**
- **Merged the data back together**
- KD, Curry, and MJ scored 30.1 points in different seasons
- But the avg and/or standard deviation was lower during Curry's season
- After standardizing data, we see Curry's points per game is more impressive than KD and MJ



PCA of NBA Players

- Performed Principal Component Analysis in R
- The first two Principal Components were able to describe at least 70% of the data which is awesome

Importance of components	Past 31 Years		Past 15 Years		Past 10 Years	
	PC1	PC2	PC1	PC2	PC1	PC2
Standard deviation	2.4746	1.1292	2.4676	1.1092	2.4654	1.0968
Proportion of Variance	0.6137	0.1278	0.6101	0.1233	0.6089	0.1205
Cumulative Proportion	0.6137	**0.7415**	0.6101	**0.7333**	0.6089	**0.7294**

Summary of PCA

- PC1 strongly describes level of responsibility
  - Higher games started(GS), more shots taken(FTA+2PA+3PA), more turnovers (TOV) and more points scored (PTS) means more responsibility
- PC2 strongly separates playstyle into guards (3PA and AST) and Big Men(TRB)(Rodrigues and Rocha da Silva)

	PTS	FTA	2PA	3PA	TS%	AST	TOV	TRB	STL	GS
31 yr PC1	0.388	0.355	0.367	0.211	0.158	0.303	0.370	0.272	0.322	0.334
31 yr PC2	0.013	0.188	0.193	-0.595	0.206	-0.407	-0.061	0.541	-0.248	0.086
15 yr PC1	0.387	0.354	0.368	0.204	0.167	0.297	0.371	0.272	0.326	0.338
15 yr PC2	0.032	0.220	0.192	-0.602	0.106	-0.428	-0.031	0.528	-0.267	0.078
10 yr PC1	0.389	0.356	0.365	0.228	0.140	0.315	0.368	0.275	0.316	0.327
10 yr PC2	0.005	-0.136	-0.184	0.544	-0.391	0.371	0.119	-0.541	0.207	-0.101

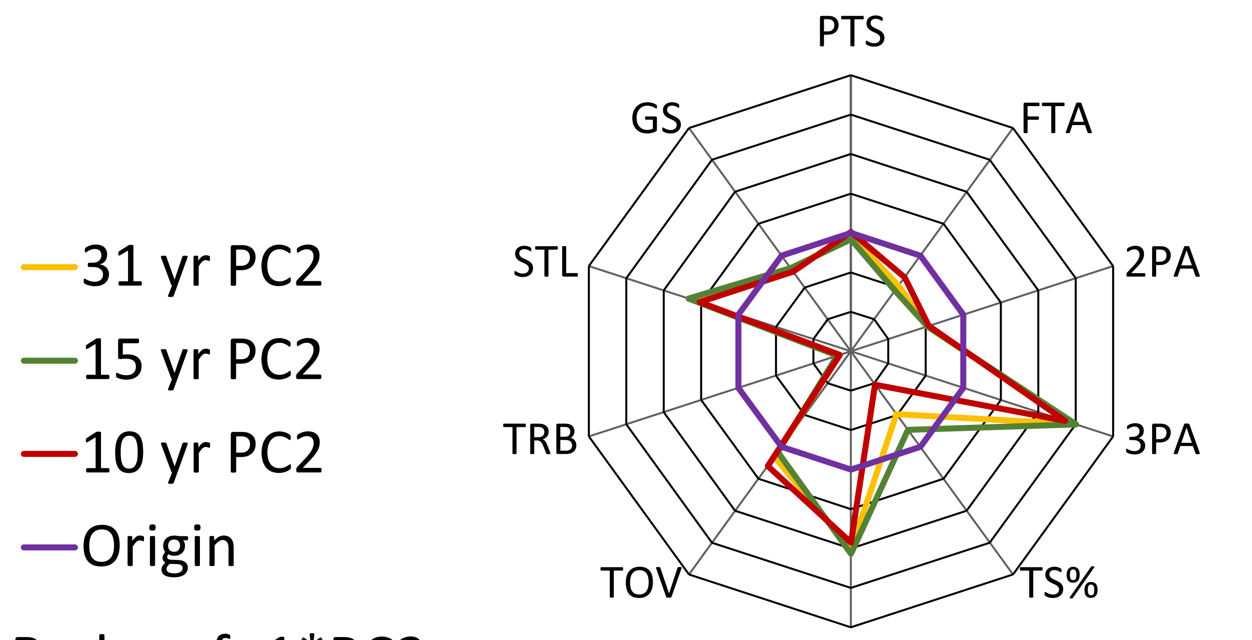
Summary of PCA

- PC1 highest magnitude:
  - PTS, TOV, 2PA, FTA ➕
- PC2 highest magnitude
  - TRB, TS% ➕
  - 3PA, AST ➖

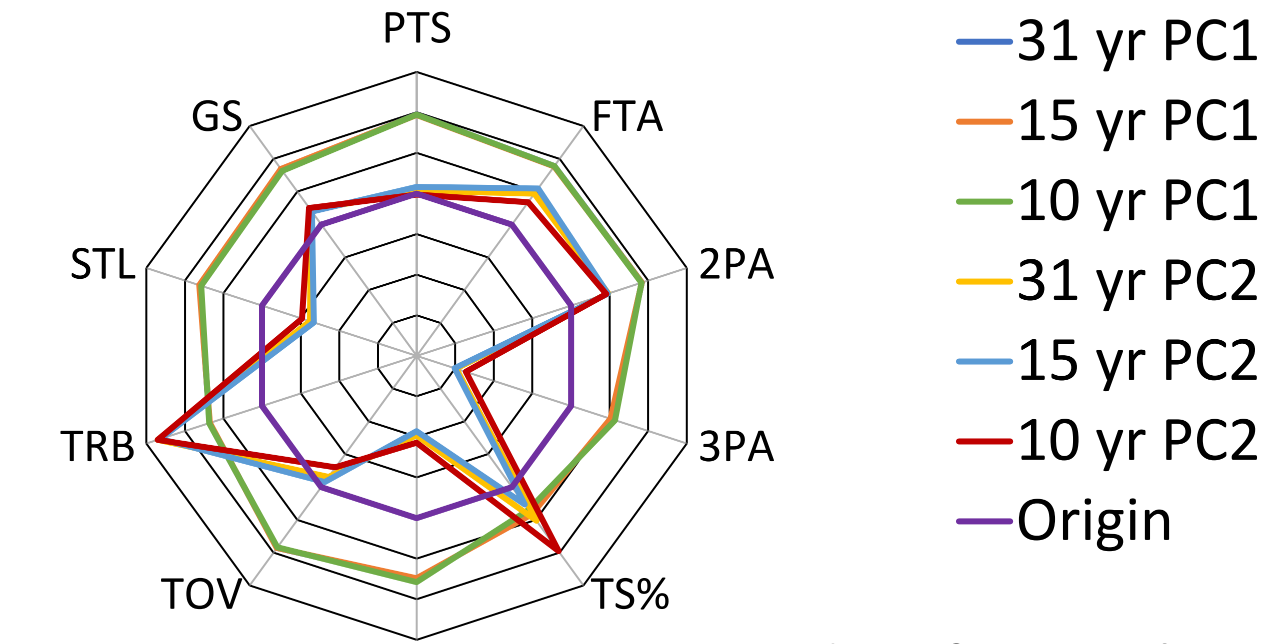
	Visually seeing Magnitudes of Principal Components									
Data	PTS	FTA	2PA	3PA	TS%	AST	TOV	TRB	STL	GS
31 yr PC1	0.388	0.355	0.367	0.211	0.158	0.303	0.370	0.272	0.322	0.334
31 yr PC2	0.013	0.188	0.193	-0.595	0.206	-0.407	-0.061	0.541	-0.248	0.086
15 yr PC1	0.387	0.354	0.368	0.204	0.167	0.297	0.371	0.272	0.326	0.338
15 yr PC2	0.032	0.220	0.192	-0.602	0.106	-0.428	-0.031	0.528	-0.267	0.078
10 yr PC1	0.389	0.356	0.365	0.228	0.140	0.315	0.368	0.275	0.316	0.327
10 yr PC2	-0.005	0.136	0.184	-0.544	0.391	-0.371	-0.119	0.541	-0.207	0.101

Radar of Principal Components

- The PC1 lines on Right Figure are stacked on top of each other (Green line)
- On Right Figure, PC2 playstyle is high rebounds and high TS% e.g Gobert
- On Left Figure, PC2 playstyle is high assists and 3p e.g Trae Young
- Vertices closer to purple origin means that stat matters less for the PC



Radar of -1*PC2

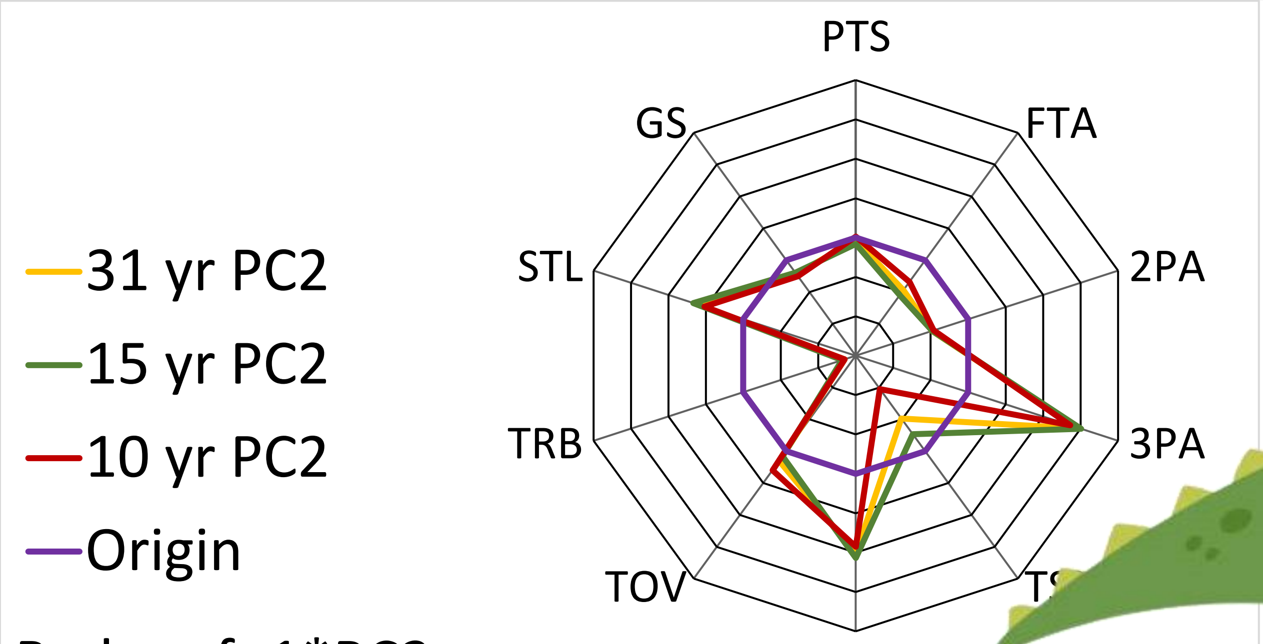


Radar of PCA Values

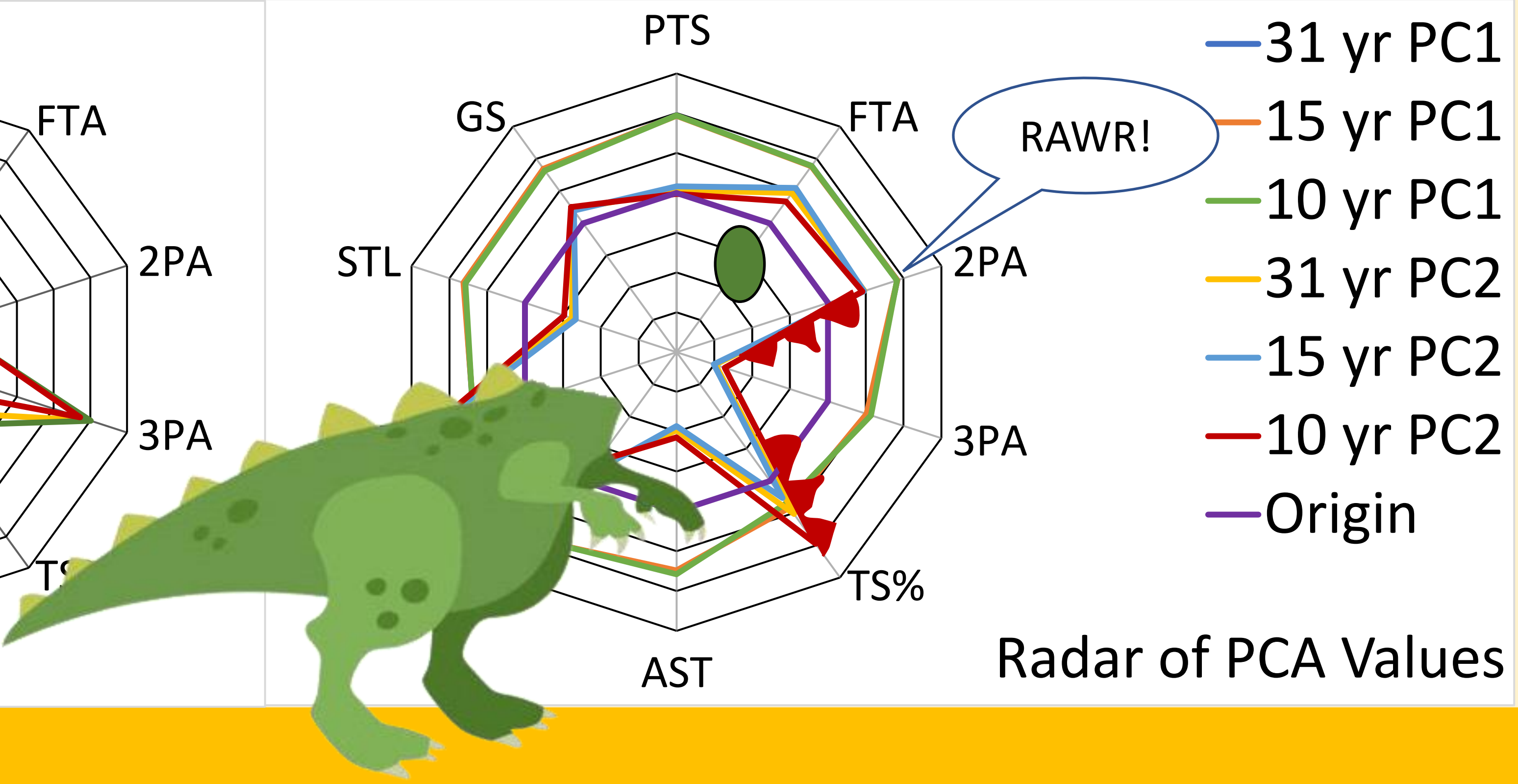
Hussaini Lomsomboot

Radar of Principal Components

- The PC1 lines on Right Figure are stacked on top of each other (Green line)
- On Right Figure, PC2 playstyle is high rebounds and high TS% e.g Gobert
- On Left Figure, PC2 playstyle is high assists and 3p e.g Trae Young
- Vertices closer to purple origin means that stat matters less for the PC



Radar of -1*PC2

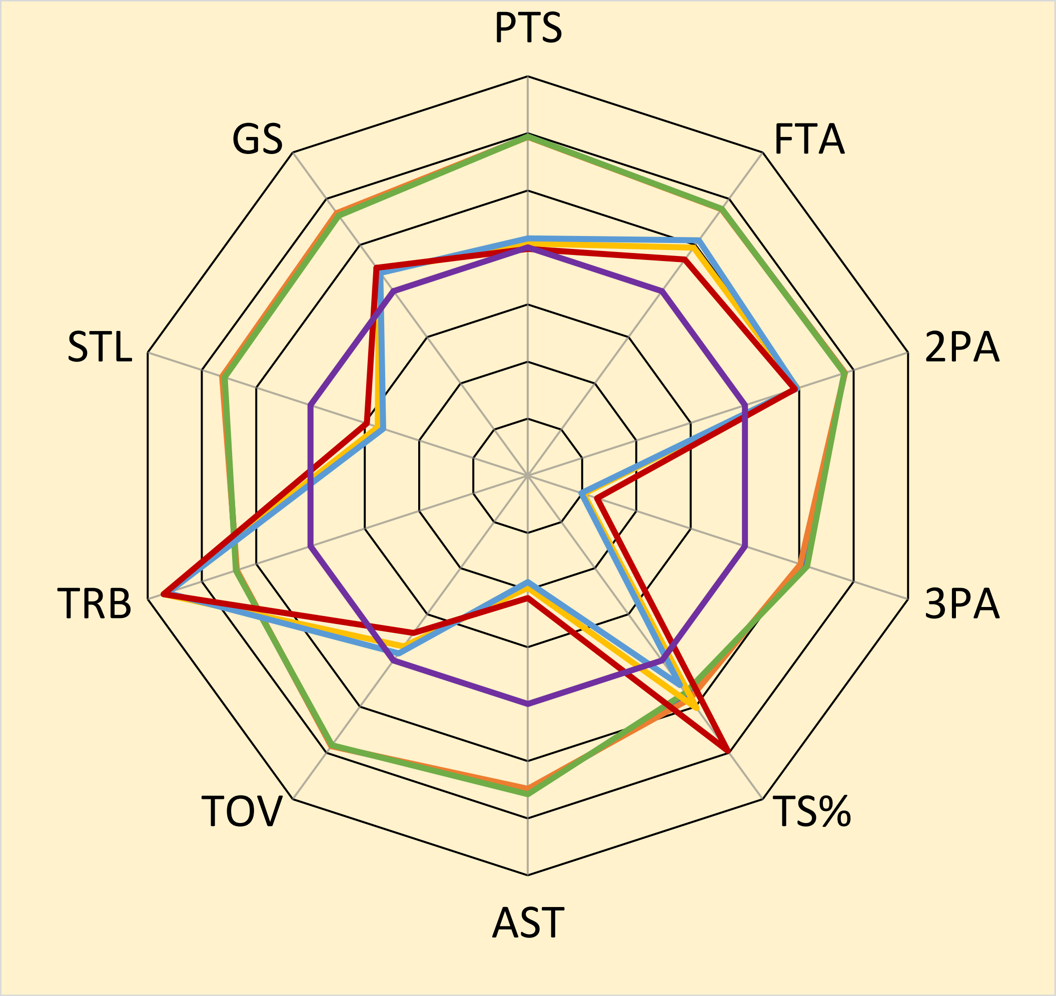


Radar of PCA Values

TS% and TRB have increased in correlation

- Big men have been forced to limit their shot selection in recent years
  - The highest % is by guards in 1995
  - The highest % is by big men in 2022

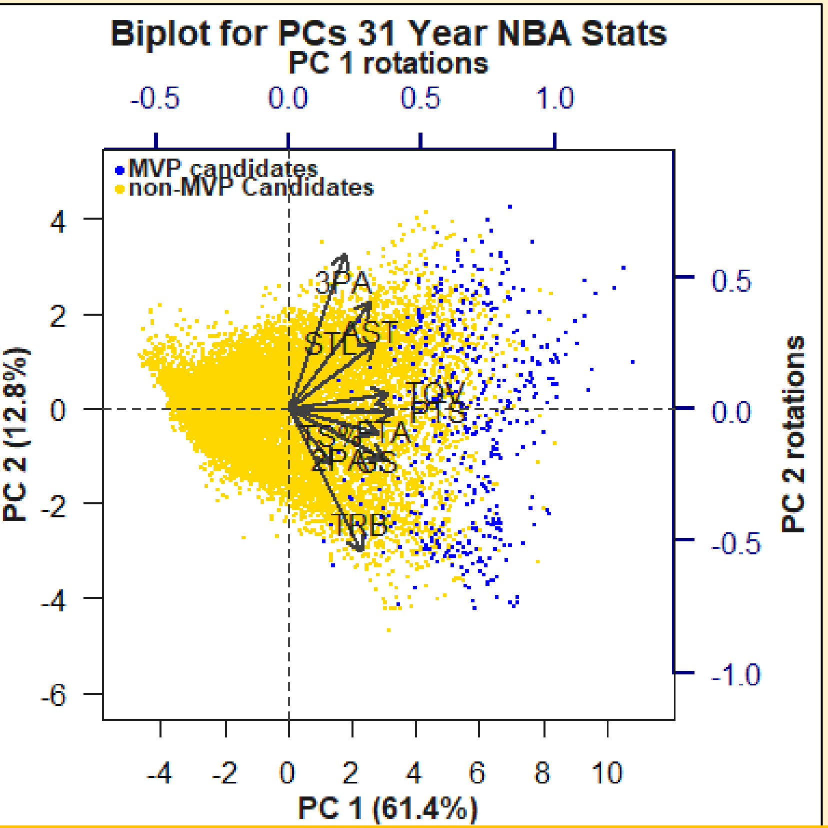
1995 players sorted by Highest TS%				2022 players sorted by Highest TS%			
Rk	**Player**	**Pos**	**TS% ▼**	**Rk**	**Player**	**Pos**	**TS% ▼**
1	John Stockton*	PG	0.65	**1**	Rudy Gobert	C	0.73
2	Detlef Schrempf	SF	0.64	**2**	Jarrett Allen	C	0.7
3	Chris Gatling	PF	0.64	**3**	Montrezl Harrell	C	0.68
4	Kenny Smith	PG	0.64	**4**	Nikola Jokić	C	0.66
5	Steve Kerr	PG	0.64	**5**	Brandon Clarke	PF	0.66
6	Dana Barros	PG	0.63	**6**	Ivica Zubac	C	0.66
7	Mario Elie	SF	0.63	**7**	Deandre Ayton	C	0.66
8	Hersey Hawkins	SG	0.62	**8**	JaVale McGee	C	0.65
9	Jeff Hornacek	SG	0.62	**9**	Karl-Anthony Towns	C	0.64
10	Reggie Miller*	SG	0.62	**10**	Domantas Sabonis	C-PF	0.64

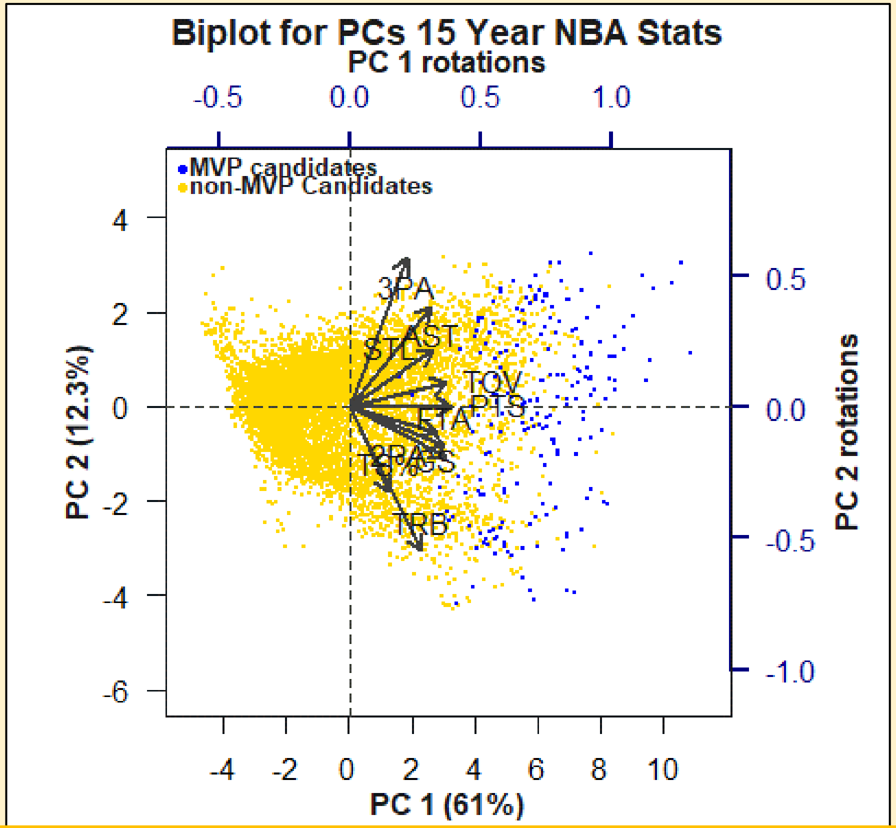


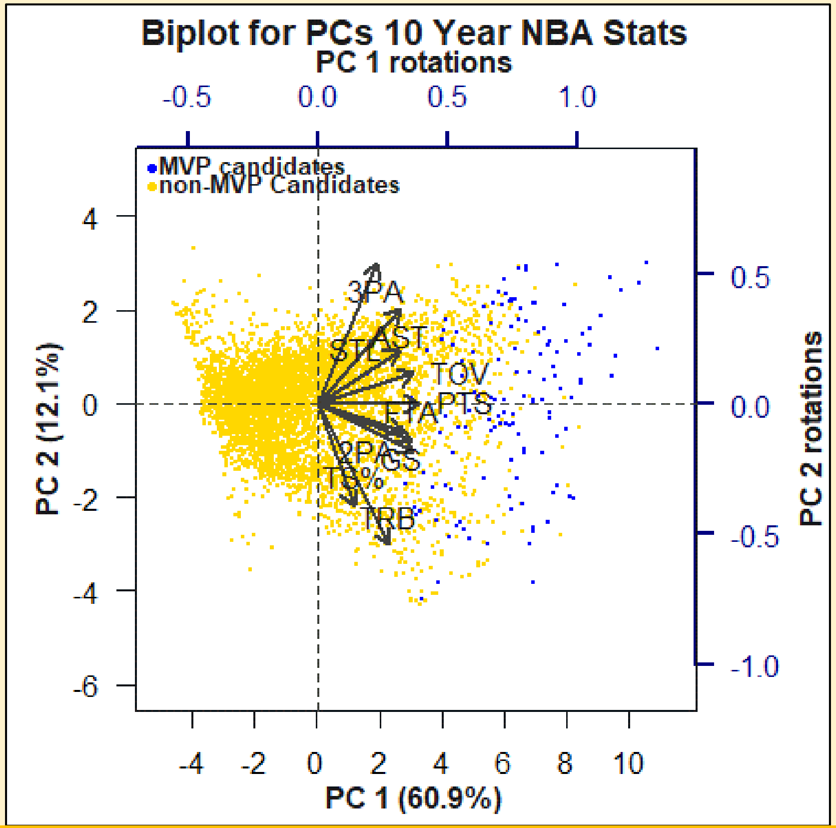
Biplots

- We can see that the MVP candidates (blue) move along the PC 1 axis
  - They are spread out along the PC 2 Axis
- Data points to the right, the higher the chance of receiving a vote

Blue dots are combination of MVP and MVP Candidates



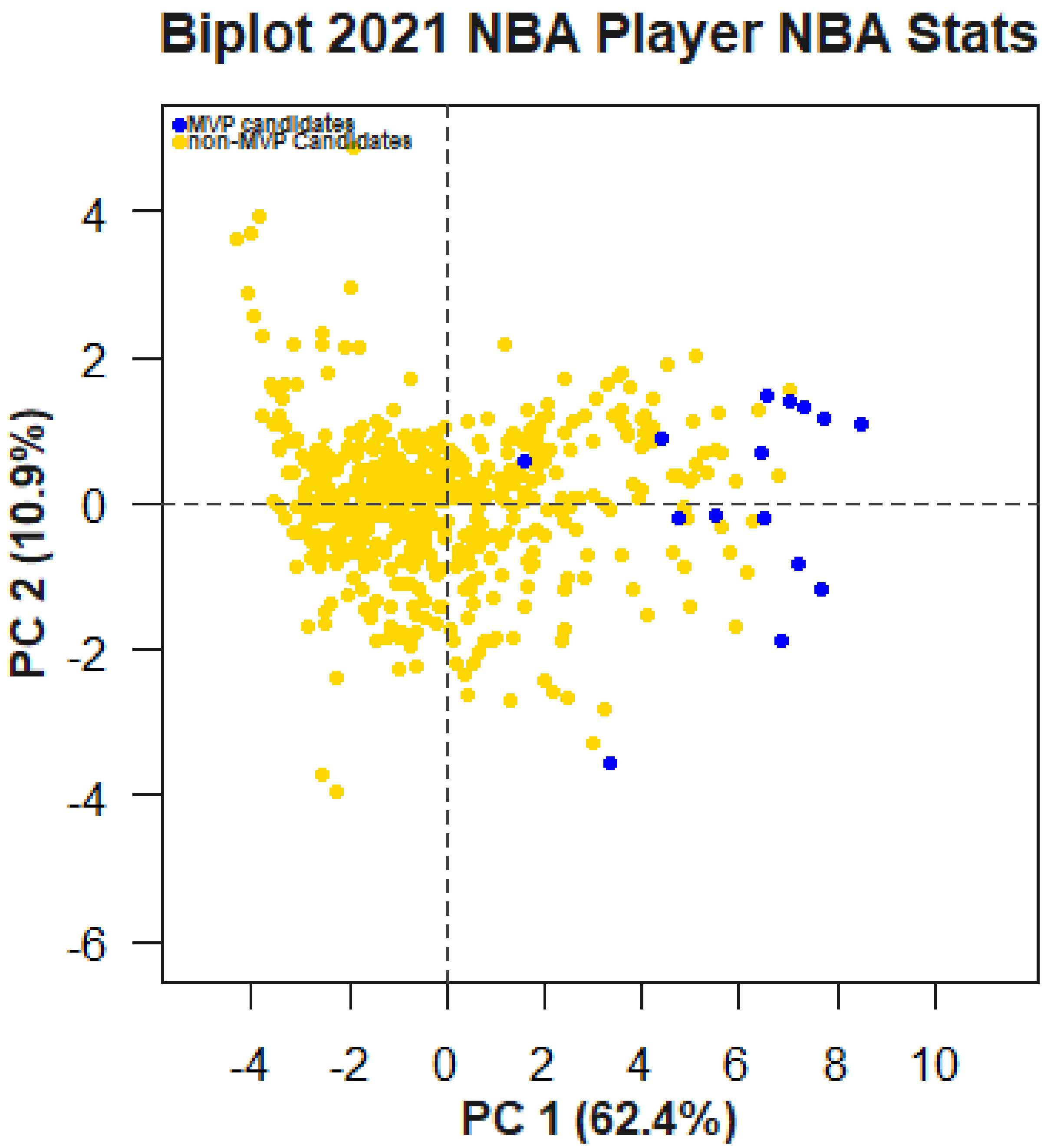




Let’s Enhance!

- Data points further to the right and further from the origin have a higher chance of receiving a vote
- Blue data point closest to origin is **Derrick Rose** who received his vote as a joke

Blue dots are combination of MVP and MVP Candidates



Determining Variable Importance with ExtraTrees

- sklearn.ensemble.ExtraTreesClassifier
  - "Ensemble of extremely randomized tree classifiers"
- Ran model 500 times and then took average of each stat
- Top five variables are
  - Points
  - Free throws attempted
  - 2 pointers attempted
  - Win % of team
  - Assists
- Package has built in prediction model

Rank	Stat	31 year	15 year	10 year	Avg
1	PTS	0.152	0.150	0.151	0.151
2	FTA	0.114	0.122	0.116	0.117
3	2PA	0.090	0.083	0.078	0.084
4	W/L%	0.085	0.082	0.077	0.081
5	AST	0.072	0.077	0.081	0.077
6	SRS	0.071	0.073	0.070	0.071
7	TOV	0.064	0.068	0.065	0.066
8	GS	0.054	0.054	0.050	0.052
9	STL	0.049	0.047	0.052	0.050
10	TRB	0.055	0.045	0.046	0.049
11	TS%	0.046	0.048	0.050	0.048
12	BLK	0.042	0.037	0.038	0.039
13	Age	0.035	0.037	0.040	0.037
14	PF	0.035	0.037	0.037	0.036
15	3PA	0.035	0.041	0.050	0.042

Determining Variable Trends with ExtraTrees

- Subtracted each row by the row average
- Importance that have steadily decreased
  - **2PA** (Superstars moving away from 2PA?)
  - **W/L%** (Does having talented teammates reduce the impact of a candidate)
  - TRB
  - GS?
- Values that have increased
  - **3PA** (Top 3 Big men in NBA can shoot 3s)
  - TS% (Do voters appreciate efficiency?)
  - Age

Rank	Stat	31 year	15 year	10 year
1	PTS	0.001	-0.001	0.000
2	FTA	-0.004	0.004	-0.001
3	2PA	0.007	-0.001	-0.006
4	W/L%	0.004	0.001	-0.005
5	AST	-0.005	0.000	0.004
6	SRS	0.000	0.002	-0.002
7	TOV	-0.002	0.002	-0.001
8	GS	0.001	0.001	-0.003
9	STL	0.000	-0.002	0.002
10	TRB	0.007	-0.004	-0.003
11	TS%	-0.002	0.000	0.002
12	BLK	0.003	-0.002	-0.001
13	Age	-0.002	0.000	0.002
14	PF	-0.002	0.001	0.001
15	3PA	-0.007	-0.001	0.008

Determining Variable Importance with Random Forest

- sklearn.ensemble.RandomForestClassifier
  - "A random forest classifier with optimal splits"
  - Ran model 100 times and took average
    Top five variables are
    - Points
    - Free throws attempted
    - Win %
    - 2 pointers attempted
    - Simple Rating System
- 3PA jumped from 15 to 12 in rank
- Package has built in prediction model

Rank	Stat	31 RF	15 RF	10 RF	Avg
1	PTS	0.182	0.176	0.191	0.183
2	FTA	0.110	0.118	0.114	0.114
3	W/L%	0.099	0.094	0.081	0.091
4	2PA	0.082	0.079	0.076	0.079
5	SRS	0.077	0.078	0.076	0.077
6	AST	0.066	0.070	0.075	0.070
7	TOV	0.066	0.067	0.062	0.065
8	GS	0.049	0.053	0.044	0.049
9	TS%	0.045	0.046	0.047	0.046
10	TRB	0.053	0.043	0.041	0.045
11	STL	0.044	0.044	0.047	0.045
12	3PA	0.030	0.037	0.046	0.038
13	Age	0.033	0.033	0.036	0.034
14	BLK	0.034	0.031	0.031	0.032
15	PF	0.032	0.032	0.032	0.032

Determining Variable Trends with Random Forest

- Subtracted each row by the row average
- Importance that have steadily decreased
  - 2PA
  - W/L%
  - TRB
- Values that have increased
  - PTS
  - BLK
  - SRS
- Similarities with Extra Trees
  - TRB 2PA W/L% TS%

Rank	Stat	31 year	15 year	10 year
1	PTS	-0.001	-0.007	0.008
2	FTA	-0.004	0.004	0.000
3	2PA	0.007	0.003	-0.010
4	W/L%	0.003	0.000	-0.003
5	AST	0.000	0.001	-0.001
6	SRS	-0.004	0.000	0.005
7	TOV	0.001	0.002	-0.003
8	GS	0.000	0.004	-0.004
9	STL	-0.001	0.000	0.001
10	TRB	0.007	-0.003	-0.005
11	TS%	-0.001	-0.001	0.002
12	BLK	-0.008	-0.001	0.008
13	Age	-0.001	-0.001	0.002
14	PF	0.002	-0.001	-0.001
15	3PA	0.000	0.000	0.000

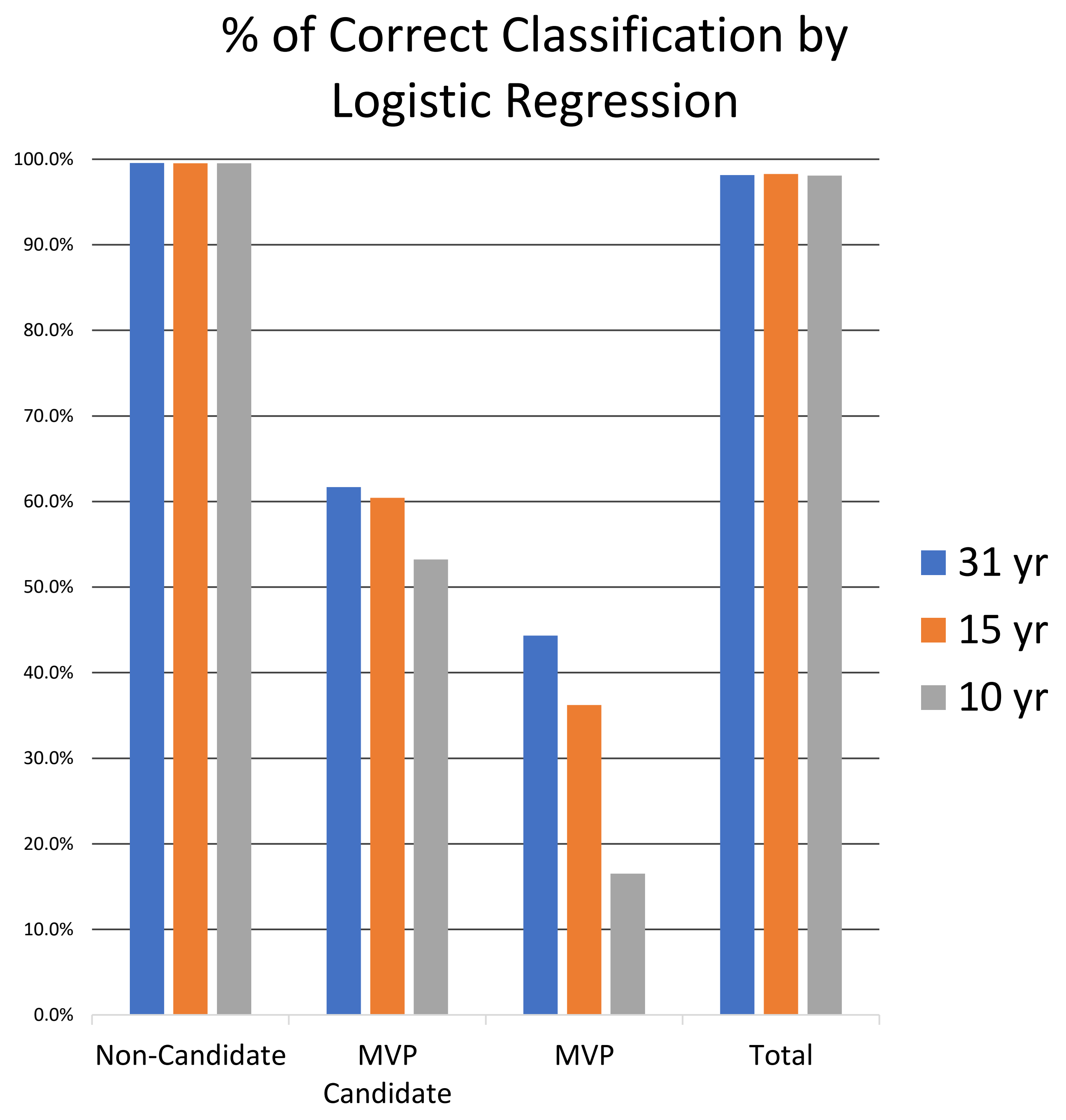
Logistic Regression Method

- sklearn.model.selection.train_test_split to randomly split data into 80% train and 20% test
- sklearn.model.selection.LogisticRegression to fit the model
- sklearn.metrics and numPY to calculate accuracy of classification of model
- model.predict() to predict MVP
- model.predict_proba() to obtain the probability of MVP prediction
- Ran the model at least a 100 times
    - Creates coefficients for variables with training data
    - Model classifies test data
        - We counted how many times it correctly and incorrectly classified an MVP, MVP Candidate, and Non-Candidate

	Non-Candidate	MVP Candidate	MVP	Total
31 yr	99.6%	61.7%	44.3%	98.1%
15 yr	99.5%	60.4%	36.2%	98.3%
10 yr	99.5%	53.2%	16.5%	98.1%

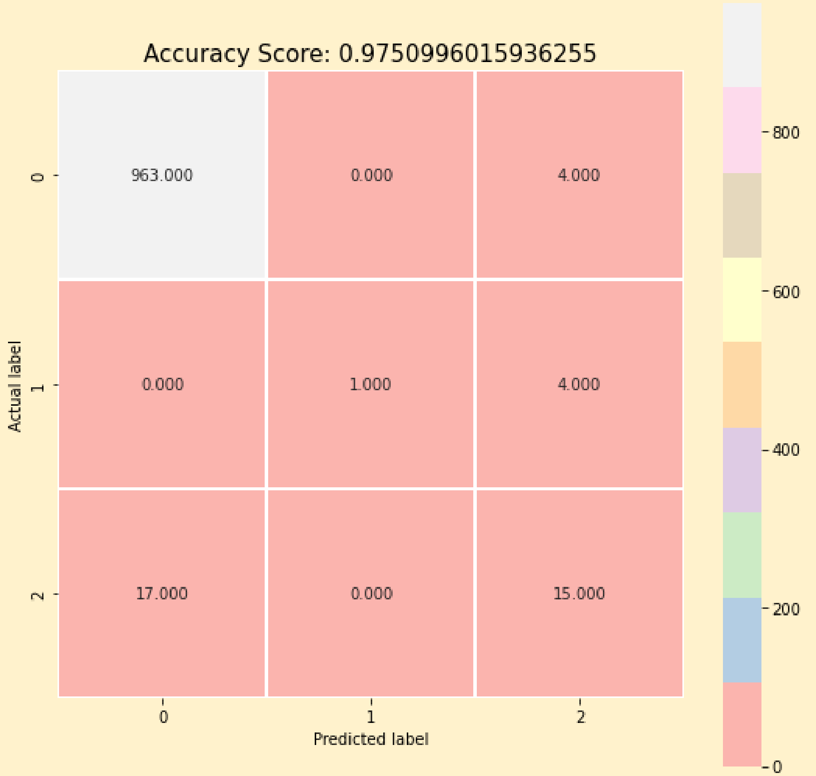
Results

- Our results had high results for Non-MVP Candidate
- 40% for MVP may be considered good but worse than real-life
- In real-life voters can make their selections public before all ballots are received
- Low accuracy for 10 years of data could be due to high variability in training data

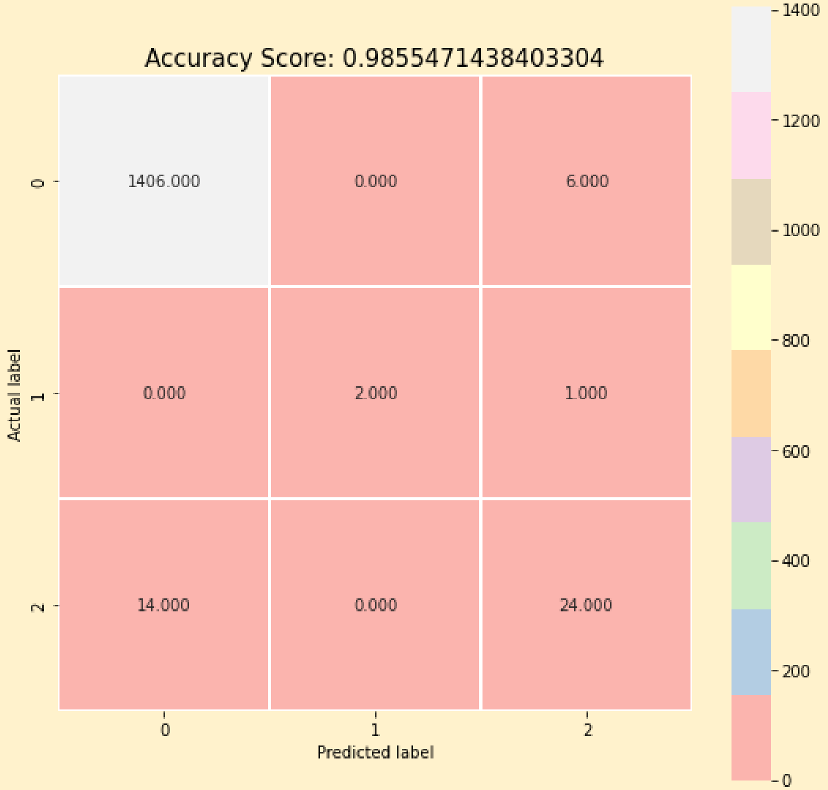


Confusion Matrix

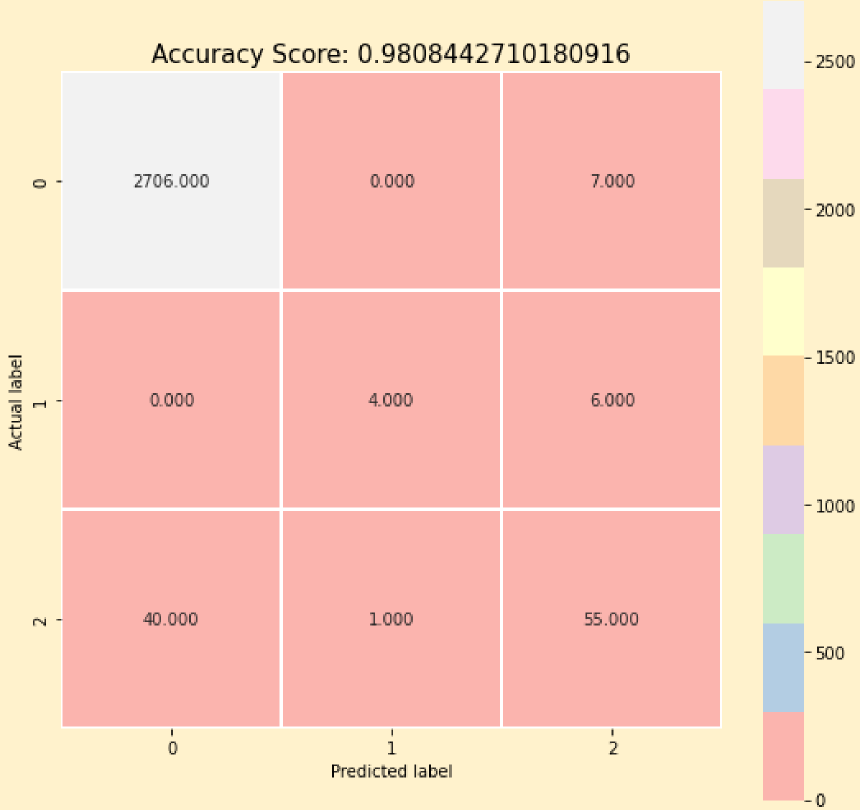
10 years



15 years



31years



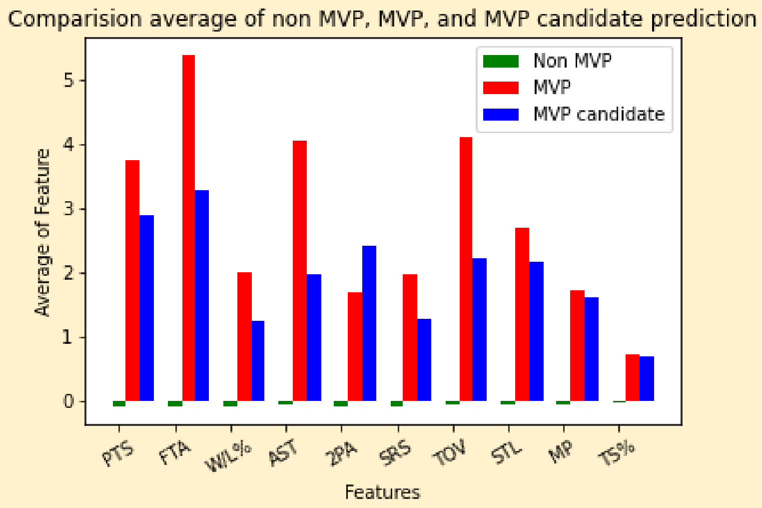
Avg. Value of Variables Trained by Logistic Regression

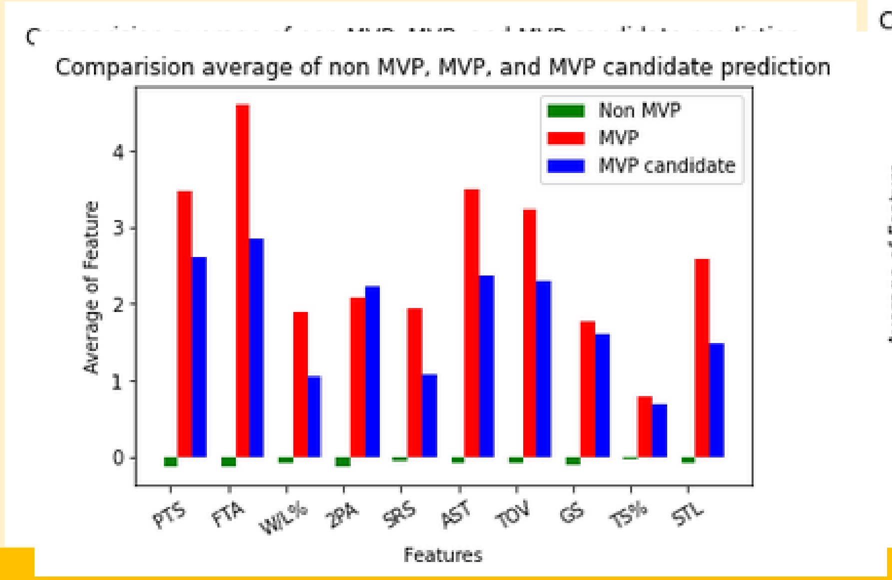
- MVPs tend to have higher values than MVP Candidates for 27/30 instances
- Non-Candidate values are negative
- 2PA on 10 years and 15 years of **MVP candidate** are higher than **MVP**

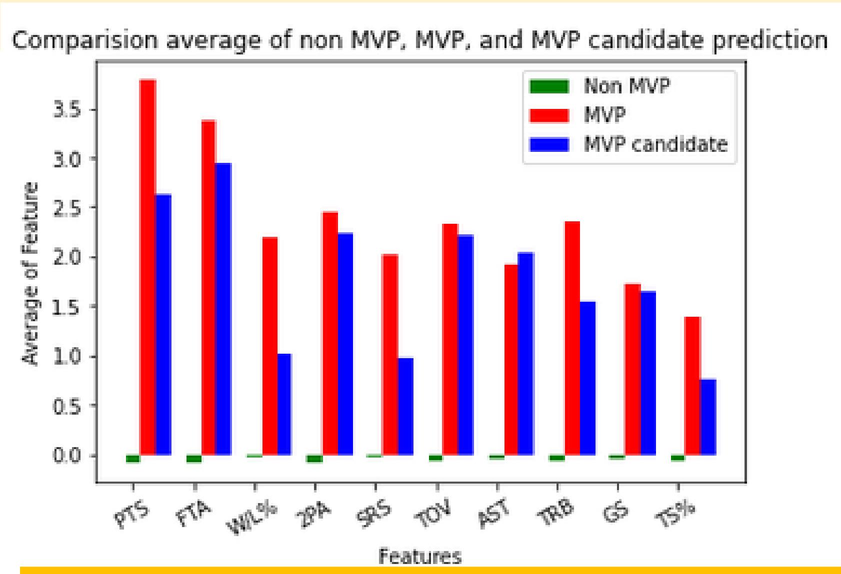
10 years

15 years

31 years







Rule based Learning to find minimum qualifications to be considered MVP and MVP candidates

Methods

- RuleFit(max_rules,rfmode= "classify", model_type="r")
- fit(X_train, y_train,features)
- get_rules()

Rule created by Rule Based – Past 10 years

PTS <= 2.314630150794983 rule -4.229046

0 : Not MVP nor MVP candidate class

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
6829	-0.554215	-1.279530	-1.379770	-0.751925	-0.867548	-0.936267	-0.555623	-0.809237	-0.923072	-1.522622	0.496858	-0.765335	-0.922376	-1.027864	-0.706557	-0.587617	-0.527129	0
6830	-1.244852	-2.044464	-1.832327	-1.016637	-0.970548	-1.168290	-1.076141	-1.179838	-1.036315	-1.522622	-0.940482	-1.021237	-1.871536	-1.391023	-0.706557	-0.587617	-5.835876	0
6831	0.366635	0.330859	-0.816833	-0.619568	-0.764549	-0.936267	-0.729129	-0.768060	-0.526721	0.425372	-0.461368	-0.509434	-0.515593	-0.978342	-0.706557	-0.587617	-0.433768	0
6832	1.517696	0.451638	-0.353238	0.174569	0.265446	-0.646239	0.080565	-0.644526	0.152739	0.181873	-0.700925	-0.253532	-0.108810	-0.285039	-0.706557	-0.587617	0.269501	0
6833	2.668758	0.008781	0.717446	0.968705	0.883443	1.151939	0.138400	1.208476	-0.186991	-0.061626	0.736415	-0.253532	0.569161	0.953004	-0.706557	-0.587617	-0.018536	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13629	1.568638	1.222186	0.534912	1.544233	1.531471	-0.484485	-0.407388	-0.808845	0.221136	-0.045857	-1.015774	-0.112113	-0.573363	0.286494	-0.268188	-0.289204	0.234289	0
13630	-0.378637	-1.095907	-1.474066	-0.997428	-1.105208	-0.836285	-0.591070	-1.268654	-0.851278	-0.561747	-0.771009	-0.709166	-0.968617	-1.024449	-0.268188	-0.289204	-0.105987	0
13631	2.055457	0.980718	0.191129	0.770684	0.696523	0.365698	-0.101252	0.486979	-0.315071	0.212088	0.452815	-0.112113	0.217146	0.379031	-0.268188	-0.289204	-0.124223	0
13632	-1.108865	-0.274916	-1.345147	-0.997428	-1.193097	-0.689702	-0.774752	-1.268654	-0.475933	-1.077636	-1.015774	-0.828577	-1.627375	-0.993603	-0.268188	-0.289204	-0.063138	0
13633	-0.135227	-0.951027	-0.453462	-0.444893	-0.489983	-0.631069	-0.468615	0.027171	-0.744037	-0.819691	-1.015774	-0.947987	-1.363872	-0.608032	-0.268188	-0.289204	0.299021	0

4885 rows × 18 columns

1 : IS MVP or MVP candidate class

0	PTS > 2.314630150794983 & W/L% > 0.8920324742794037	rule	1.714740	(

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
6919	0.136422	0.894494	1.534256	3.218760	3.252431	1.790002	3.839859	-0.356281	2.304360	0.668871	-0.461368	2.177533	0.704756	3.379568	1.101833	0.567176	1.083640	1
7007	-0.784427	0.854235	1.788130	-0.354856	-0.198051	4.023223	3.897694	3.390901	0.152739	1.642869	4.329765	1.665729	0.704756	3.231003	-0.610208	-0.384683	0.605381	1
7048	0.366635	0.330859	1.490104	1.498130	1.449940	1.964019	2.509647	1.949677	1.681522	1.155870	2.892425	1.409828	0.297973	2.752293	2.369188	2.765632	1.396498	1
7054	0.366635	1.015273	1.490104	4.409966	4.024928	1.035928	1.584282	0.384919	2.700711	2.860365	-0.461368	2.433434	0.840350	2.785308	2.369188	2.765632	1.061002	1
7080	0.366635	1.095793	1.622560	2.292267	2.582935	3.501171	4.938730	2.937944	4.852332	2.373366	0.017745	5.504253	0.840350	3.825263	0.560798	0.299013	0.290828	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13400	-0.378637	0.739250	1.351394	-0.997428	-1.149153	1.626315	1.919248	1.490198	2.634066	2.533592	0.452815	2.276098	1.666412	0.826294	1.321222	1.129245	0.370501	1
13408	0.108182	0.401195	1.211733	0.107642	0.081298	3.004199	5.470428	2.911425	0.435618	0.985923	2.410934	2.395509	1.007655	3.016339	1.321222	1.129245	0.909177	1
13444	0.108182	0.884131	1.415854	0.107642	0.344966	2.945566	4.735701	3.078628	2.097860	1.501812	1.921404	2.753741	1.534661	2.954648	1.021871	1.189209	0.857341	1
13482	-1.108865	1.125599	1.555515	2.096768	2.410364	2.300599	3.266247	1.824605	3.545618	0.985923	0.208050	3.828436	0.875903	2.892957	0.622737	0.504797	0.398510	1
13522	0.595001	1.367067	1.179503	-1.107935	-1.193097	1.098615	2.164157	4.123648	-0.368692	-0.045857	5.592878	0.723761	0.875903	0.826294	1.613446	1.892230	1.299101	1

133 rows × 18 columns

Rule created by Rule Based – Past 15 years

FTA <= 2.8089007139205933 rule -4.132595

0 : Not MVP nor MVP candidate class

class_0

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
6829	-0.554215	-1.279530	-1.379770	-0.751925	-0.867548	-0.936267	-0.555623	-0.809237	-0.923072	-1.522622	0.496858	-0.765335	-0.922376	-1.027864	-0.706557	-0.587617	-0.527129	0
6830	-1.244852	-2.044464	-1.832327	-1.016637	-0.970548	-1.168290	-1.076141	-1.179838	-1.036315	-1.522622	-0.940482	-1.021237	-1.871536	-1.391023	-0.706557	-0.587617	-5.835876	0
6831	0.366635	0.330859	-0.816833	-0.619568	-0.764549	-0.936267	-0.729129	-0.768060	-0.526721	0.425372	-0.461368	-0.509434	-0.515593	-0.978342	-0.706557	-0.587617	-0.433768	0
6832	1.517696	0.451638	-0.353238	0.174569	0.265446	-0.646239	0.080565	-0.644526	0.152739	0.181873	-0.700925	-0.253532	-0.108810	-0.285039	-0.706557	-0.587617	0.269501	0
6833	2.668758	0.008781	0.717446	0.968705	0.883443	1.151939	0.138400	1.208476	-0.186991	-0.061626	0.736415	-0.253532	0.569161	0.953004	-0.706557	-0.587617	-0.018536	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14087	-0.091724	1.113670	1.025399	0.617089	0.763783	0.587060	0.557555	0.723273	0.570653	0.877153	-0.182383	0.660981	0.628128	0.730006	-0.537997	-0.676847	0.302459	0
14088	-0.555920	-0.563355	-1.363893	-0.774400	-0.660391	-1.038473	-0.818088	-0.967234	-0.698884	-0.495522	-0.585917	-0.663611	-1.272335	-1.066681	-0.537997	-0.676847	-0.908281	0
14089	-0.555920	-0.604258	-1.099466	-0.619790	-0.541710	-0.886079	-0.563339	-1.128235	-0.809279	-0.724301	-0.585917	-0.904446	-1.628672	-0.907682	-0.537997	-0.676847	-0.480678	0
14090	-1.484312	-0.358839	-0.806706	-0.155960	-0.185666	-0.759085	-0.919988	-0.685483	-0.698884	-1.181859	-0.384150	-0.904446	-0.559662	-0.685084	-0.537997	-0.676847	0.202020	0
14091	-0.091724	0.541028	0.619314	0.617089	0.467080	0.282272	-0.053842	1.045274	0.073878	0.190816	0.422918	-0.181941	1.103244	0.332509	-0.537997	-0.676847	0.370827	0

7059 rows × 18 columns

1 : IS MVP or MVP candidate class

FTA > 2.8089007139205933 rule 0.097159

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
6919	0.136422	0.894494	1.534256	3.218760	3.252431	1.790002	3.839859	-0.356281	2.304360	0.668871	-0.461368	2.177533	0.704756	3.379568	1.101833	0.567176	1.083640	1
7007	-0.784427	0.854235	1.788130	-0.354856	-0.198051	4.023223	3.897694	3.390901	0.152739	1.642869	4.329765	1.665729	0.704756	3.231003	-0.610208	-0.384683	0.605381	1
7048	0.366635	0.330859	1.490104	1.498130	1.449940	1.964019	2.509647	1.949677	1.681522	1.155870	2.892425	1.409828	0.297973	2.752293	2.369188	2.765632	1.396498	1
7054	0.366635	1.015273	1.490104	4.409966	4.024928	1.035928	1.584282	0.384919	2.700711	2.860365	-0.461368	2.433434	0.840350	2.785308	2.369188	2.765632	1.061002	1
7080	0.366635	1.095793	1.622560	2.292267	2.582935	3.501171	4.938730	2.937944	4.852332	2.373366	0.017745	5.504253	0.840350	3.825263	0.560798	0.299013	0.290828	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13949	0.372472	0.950058	1.469260	0.617089	0.467080	2.466582	2.493645	2.172279	0.901836	0.190816	0.826451	1.022234	0.271791	2.606192	2.476057	2.000638	0.989980	1
13972	0.836668	1.031864	1.270939	-0.774400	-0.779072	2.212592	2.493645	2.856531	0.901836	0.419595	4.054722	1.865156	0.628128	1.874797	1.625939	2.290023	0.730476	1
13973	-0.555920	0.909155	1.119838	-0.465180	-0.541710	2.136395	1.168951	-0.121981	2.060979	1.105932	-0.585917	1.503904	-0.203325	1.652199	1.625939	2.290023	0.669302	1
14015	0.140374	0.663737	1.431485	2.008578	2.365979	2.593576	2.391745	0.723273	2.612951	1.563491	0.221151	2.105991	-0.084546	2.606192	1.061770	1.394824	0.049740	1
14022	0.836668	0.868252	1.771463	-0.465180	-0.423028	2.949162	2.238896	3.742035	1.288217	1.334712	2.642354	1.744739	0.509349	2.256395	-0.823946	-0.822892	0.368273	1

204 rows × 18 columns

Rule created by Rule Based – Past 31 years

1	PTS <= 1.667323112487793	rule	0.0	0.918589	0.0

0 : Not MVP nor MVP Candidate class

```
] class_0 = df.iloc[np.where(df.Is_MVP_or_MVP_candidate == "0")]
  class_0
```

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
0	-0.886218	-0.716531	-1.501674	-0.726261	-0.796777	-1.057710	-0.606015	-0.758695	-0.969130	-1.214828	-0.420182	-1.004570	-1.330463	-1.064343	0.472185	0.509455	0.021818	0
1	0.174886	-1.903423	-1.558538	-0.726261	-0.796777	-1.258329	-0.771329	-0.796254	-1.020713	-1.214828	-0.786541	-1.483782	-1.450498	-1.277519	0.472185	0.509455	-1.687118	0
2	0.174886	0.232983	0.100016	-0.726261	-0.796777	-0.079690	-0.220282	0.405639	-0.659630	-0.011881	0.312536	-0.165949	0.109958	-0.277230	0.472185	0.509455	0.293066	0
3	2.562368	-2.061675	-0.828774	-0.726261	-0.796777	-1.158019	-0.881538	-0.458222	-1.020713	-1.415319	-0.237003	-1.603585	0.710134	-1.228325	0.472185	0.509455	-1.371853	0
4	-0.355666	-0.558279	-0.080055	-0.726261	-0.735300	-0.355542	-0.440701	0.405639	-0.659630	-0.813845	-0.237003	-0.764964	0.590099	-0.392018	0.472185	0.509455	1.134430	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14087	-0.091724	1.113670	1.025399	0.617089	0.763783	0.587060	0.557555	0.723273	0.570653	0.877153	-0.182383	0.660981	0.628128	0.730006	-0.537997	-0.676847	0.302459	0
14088	-0.555920	-0.563355	-1.363893	-0.774400	-0.660391	-1.038473	-0.818088	-0.967234	-0.698884	-0.495522	-0.585917	-0.663611	-1.272335	-1.066681	-0.537997	-0.676847	-0.908281	0
14089	-0.555920	-0.604258	-1.099466	-0.619790	-0.541710	-0.886079	-0.563339	-1.128235	-0.809279	-0.724301	-0.585917	-0.904446	-1.628672	-0.907682	-0.537997	-0.676847	-0.480678	0
14090	-1.484312	-0.358839	-0.806706	-0.155960	-0.185666	-0.759085	-0.919988	-0.685483	-0.698884	-1.181859	-0.384150	-0.904446	-0.559662	-0.685084	-0.537997	-0.676847	0.202020	0
14091	-0.091724	0.541028	0.619314	0.617089	0.467080	0.282272	-0.053842	1.045274	0.073878	0.190816	0.422918	-0.181941	1.103244	0.332509	-0.537997	-0.676847	0.370827	0

13621 rows × 18 columns

1 : IS MVP or MVP candidate class

	rule	type	coef	support	importance
0	PTS > 1.667323112487793	rule	0.0	0.081411	0.0

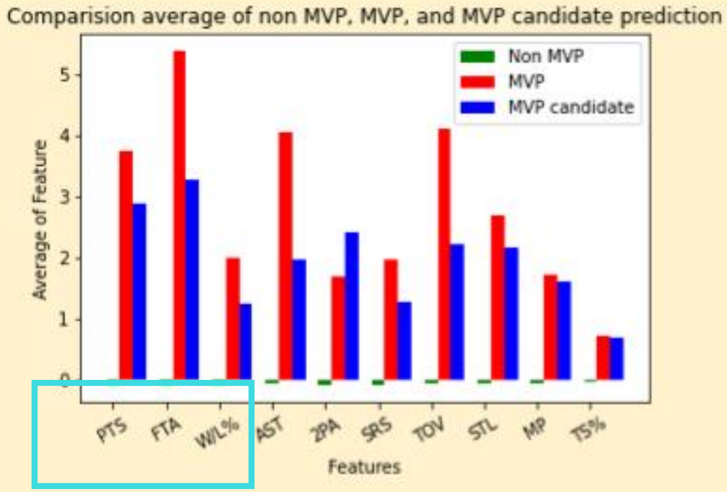
```
] class_1 = df.iloc[np.where(df.Is_MVP_or_MVP_candidate == "1")]
  class_1
```

	Age	G	MP	3P	3PA	2PA	FTA	TRB	AST	STL	BLK	TOV	PF	PTS	W/L%	SRS	TS%	Is_MVP_or_MVP_candidate
5	1.235989	0.826429	1.436338	2.329278	2.523001	1.324645	1.928798	1.044144	1.455287	2.193521	0.312536	1.271686	0.710134	2.182496	0.472185	0.509455	0.629882	1
6	1.235989	0.668176	1.787004	-0.726261	-0.673822	3.907619	2.865577	2.734306	0.784703	2.193521	5.441561	2.349913	1.670415	3.166386	0.472185	0.509455	0.449002	1
18	0.970713	0.509924	1.351041	1.042735	1.170499	2.001735	2.865577	2.846983	1.094203	1.792539	0.495715	1.032080	1.070239	2.379274	1.394055	0.837225	0.568083	1
83	1.235989	1.063807	1.351041	1.203553	0.924589	0.321548	0.771601	-0.157748	5.324037	3.396468	-0.237003	2.349913	0.590099	1.018225	1.469310	1.667292	1.410709	1
84	0.970713	1.063807	1.644842	-0.565443	-0.489390	3.205451	3.416624	2.659187	0.784703	1.792539	1.045254	1.870701	1.430344	2.986006	1.469310	1.667292	0.747133	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13949	0.372472	0.950058	1.469260	0.617089	0.467080	2.466582	2.493645	2.172279	0.901836	0.190816	0.826451	1.022234	0.271791	2.606192	2.476057	2.000638	0.989980	1
13972	0.836668	1.031864	1.270939	-0.774400	-0.779072	2.212592	2.493645	2.856531	0.901836	0.419595	4.054722	1.865156	0.628128	1.874797	1.625939	2.290023	0.730476	1
13973	-0.555920	0.909155	1.119838	-0.465180	-0.541710	2.136395	1.168951	-0.121981	2.060979	1.105932	-0.585917	1.503904	-0.203325	1.652199	1.625939	2.290023	0.669302	1
14015	0.140374	0.663737	1.431485	2.008578	2.365979	2.593576	2.391745	0.723273	2.612951	1.563491	0.221151	2.105991	-0.084546	2.606192	1.061770	1.394824	0.049740	1
14022	0.836668	0.868252	1.771463	-0.465180	-0.423028	2.949162	2.238896	3.742035	1.288217	1.334712	2.642354	1.744739	0.509349	2.256395	-0.823946	-0.822892	0.368273	1

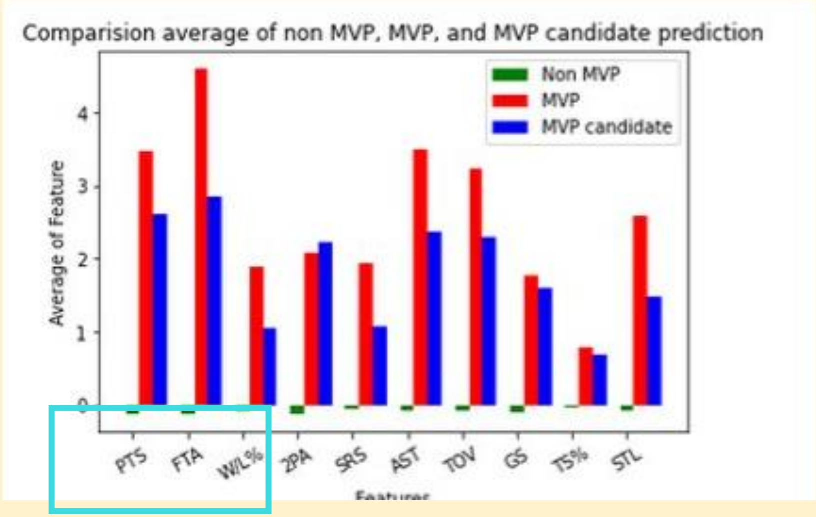
471 rows × 18 columns

Expectedly the 3 variables created by RuleFit are the top 3 from the most important variables by ExtraTreesClassifier()

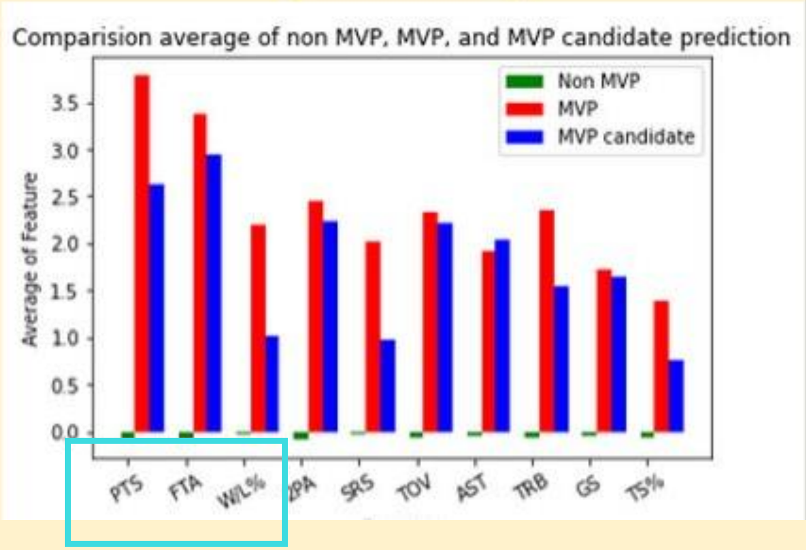
10 years



15 years



31 years



PTS
FTA
W/L%

From RuleFit

MVP Candidate Prediction

- These are the added percentages of a player being classified as an MVP or MVP Candidate
- Top 6 is accurate combination of players except for Ja Morant

MVP Candidate Percentage							
AVG	**10 yr**	**15 yr**	**31 yr**	**10 yr**	**15 yr**	**31 yr**	**Player (Team)**
99.2%	98.4%	99.4%	99.8%	98.4%	99.4%	99.8%	Giannis Antetokounmpo
98.6%	97.2%	99.2%	99.6%	97.1%	99.2%	99.7%	Joel Embiid (76ers)
96.4%	94.2%	95.5%	99.8%	94.0%	95.4%	99.7%	Nikola Jokić
96.2%	95.4%	97.3%	96.1%	95.3%	97.2%	95.9%	Ja Morant
93.3%	88.0%	95.7%	95.8%	88.0%	95.5%	96.6%	Devin Booker (Suns)
93.3%	86.7%	94.0%	99.3%	86.0%	94.0%	99.5%	Luka Dončić
89.1%	83.0%	92.6%	92.7%	82.4%	92.7%	91.0%	Trae Young
88.4%	86.3%	86.0%	94.5%	86.1%	85.8%	92.0%	Kevin Durant
85.4%	82.8%	95.5%	81.1%	83.4%	95.2%	74.6%	DeMar DeRozan
82.0%	68.1%	81.7%	96.1%	66.9%	82.1%	97.1%	James Harden (76ers)
81.0%	77.2%	77.5%	87.8%	77.3%	77.0%	88.9%	Chris Paul (Suns)
75.7%	76.7%	81.2%	66.5%	77.0%	80.8%	71.9%	Jimmy Butler
71.3%	51.0%	75.5%	86.1%	50.0%	75.1%	90.4%	Jayson Tatum
52.0%	35.1%	40.3%	78.2%	34.3%	40.2%	84.0%	Stephen Curry
53.9%	39.2%	58.4%	58.4%	39.5%	57.5%	70.5%	Pascal Siakam

MVP Prediction

- Believe algorithm did a good job considering that it does not have access to the internet and public opinions
- Most of the high probabilities are near the top
  - Algorithms can predict top 5 well
- **Ja Morant** won Most Improved Player Award so voters may not vote for him for MVP

Logisitic Regression

Random Forest Classifier			Extra Trees Classifier			Basketball Reference Algorithm		Votes Counted as of 5/06					
10 yr Prob%	15 yr Prob%	31 yr Prob%	10 yr Prob%	15 yr Prob%	31 yr Prob%	BR Prob%	Player (Team)	1st Place Votes	2nd place votes	3rd place votes	4th place votes	5th place votes	TOTAL POINTS
7.3%	5.4%	30.1%	7.1%	5.3%	23.0%	43.5%	Nikola Jokić	37	10	1	0	0	445
9.8%	15.8%	15.3%	10.4%	16.2%	18.5%	12.4%	Joel Embiid (76ers)	11	15	18	0	0	305
13.1%	19.9%	22.0%	13.8%	20.2%	23.4%	24.3%	Giannis Antetokounmpo	6	15	17	0	0	250
7.5%	21.2%	21.0%	7.5%	21.5%	31.4%	2.2%	Devin Booker (Suns)	0	1	1	17	5	68
5.8%	7.1%	11.7%	6.0%	6.9%	18.9%	4.5%	Luka Dončić	0	1	0	10	15	52
0.8%	0.7%	1.2%	1.0%	0.7%	2.3%	2.5%	James Harden (76ers)	54	42	37	27	20	1120
2.1%	2.0%	4.9%	2.1%	2.0%	7.9%	5.4%	Chris Paul (Suns)	Tabe above curated by Max Croes, @CroesFire, /u/TexasAlaskaMontana					
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	Rudy Gobert ←	This player is considered one of the best defensive players in the NBA					
2.1%	3.5%	2.0%	2.0%	3.3%	2.1%	1.6%	Trae Young						
0.0%	1.7%	1.9%	1.2%	0.6%	4.1%	1.5%	Jayson Tatum ←	Won 4th and 5th place votes			4	7	19
15.5%	20.3%	15.0%	15.9%	20.5%	16.1%	N/A	Ja Morant ←	Team won 80% of games without Ja as opposed to 63% when he played					
2.5%	5.3%	2.7%	2.5%	5.0%	1.9%	N/A	DeMar DeRozan ←	Was considered top 5 MVP for 1st half of season					

MVP Prediction

- Lack of **Jayson Tatum** is a shocker as he may end up 6th place
- Algorithm does not value defense as Rudy Gobert had 0%
  - To be fair he didn't even win Defensive Player of the Year in 2022
- Surprised **DeMar DeRozan** is that high
- We wish there was stats about percentage of points scored from isolation

Logisitic Regression

Random Forest Classifier			Extra Trees Classifier			Basketball Reference Algorithm		Votes Counted as of 5/06					
10 yr Prob%	15 yr Prob%	31 yr Prob%	10 yr Prob%	15 yr Prob%	31 yr Prob%	BR Prob%	Player (Team)	1st Place Votes	2nd place votes	3rd place votes	4th place votes	5th place votes	TOTAL POINTS
7.3%	5.4%	30.1%	7.1%	5.3%	23.0%	43.5%	Nikola Jokić	37	10	1	0	0	445
9.8%	15.8%	15.3%	10.4%	16.2%	18.5%	12.4%	Joel Embiid (76ers)	11	15	18	0	0	305
13.1%	19.9%	22.0%	13.8%	20.2%	23.4%	24.3%	Giannis Antetokounmpo	6	15	17	0	0	250
7.5%	21.2%	21.0%	7.5%	21.5%	31.4%	2.2%	Devin Booker (Suns)	0	1	1	17	5	68
5.8%	7.1%	11.7%	6.0%	6.9%	18.9%	4.5%	Luka Dončić	0	1	0	10	15	52
0.8%	0.7%	1.2%	1.0%	0.7%	2.3%	2.5%	James Harden (76ers)	54	42	37	27	20	1120
2.1%	2.0%	4.9%	2.1%	2.0%	7.9%	5.4%	Chris Paul (Suns)	Tabe above curated by Max Croes, @CroesFire, /u/TexasAlaskaMontana					
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	Rudy Gobert ←	This player is considered one of the best defensive players in the NBA					
2.1%	3.5%	2.0%	2.0%	3.3%	2.1%	1.6%	Trae Young						
0.0%	1.7%	1.9%	1.2%	0.6%	4.1%	1.5%	Jayson Tatum ←	Won 4th and 5th place votes			4	7	19
15.5%	20.3%	15.0%	15.9%	20.5%	16.1%	N/A	Ja Morant ←	Team won 80% of games without Ja as opposed to 63% when he played					
2.5%	5.3%	2.7%	2.5%	5.0%	1.9%	N/A	DeMar DeRozan ←	Was considered top 5 MVP for 1st half of season					

Algorithms vs. Official Results

Logisitic Regression													
Random Forest Classifier			Extra Trees Classifier			Basketball Reference Algorithm		Official Vote Count released May 11th					
10 yr Prob%	15 yr Prob%	31 yr Prob%	10 yr Prob%	15 yr Prob%	31 yr Prob%	BR Prob%	Player (Team)	TOTAL POINTS	1st Place Votes	2nd place votes	3rd place votes	4th place votes	5th place votes
7.3%	5.4%	30.1%	7.1%	5.3%	23.0%	43.5%	Nikola Jokić	875	65	27	6	2	0
9.8%	15.8%	15.3%	10.4%	16.2%	18.5%	12.4%	Joel Embiid (76ers)	706	26	39	34	1	0
13.1%	19.9%	22.0%	13.8%	20.2%	23.4%	24.3%	Giannis Antetokounmpo	595	9	32	52	7	0
7.5%	21.2%	21.0%	7.5%	21.5%	31.4%	2.2%	Devin Booker (Suns)	216	0	1	8	49	22
5.8%	7.1%	11.7%	6.0%	6.9%	18.9%	4.5%	Luka Dončić	146	0	1	0	32	43
0.0%	1.7%	1.9%	1.2%	0.6%	4.1%	1.5%	Jayson Tatum	43	0	0	0	8	19
15.5%	20.3%	15.0%	15.9%	20.5%	16.1%	0.0%	Ja Morant	10	0	0	0	1	7
0.2%	0.2%	0.4%	0.2%	0.2%	0.9%	0.0%	Steph Curry	4	0	0	0	0	4
2.1%	2.0%	4.9%	2.1%	2.0%	7.9%	5.4%	Chris Paul (Suns)	2	0	0	0	0	2
2.5%	5.3%	2.7%	2.5%	5.0%	1.9%	0.0%	DeMar DeRozan	1	0	0	0	0	1
0.1%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	Lebron James	1	0	0	0	0	1
0.0%	2.2%	1.8%	0.0%	2.2%	1.3%	0.0%	Kevin Durant	1	0	0	0	0	1
0.8%	0.7%	1.2%	1.0%	0.7%	2.3%	2.5%	James Harden (76ers)	0	0	0	0	0	0
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	Rudy Gobert	0	0	0	0	0	0
2.1%	3.5%	2.0%	2.0%	3.3%	2.1%	1.6%	Trae Young	0	0	0	0	0	0

Basketball Reference vs. Our Algorithm

31 RF	BR	Player
23.8%	43.5%	Nikola Jokić
12.1%	12.4%	Joel Embiid (76ers)
17.4%	24.3%	Giannis Antetokounmpo
16.6%	2.2%	Devin Booker (Suns)
9.2%	4.5%	Luka Dončić
N/A	2.5%	James Harden (76ers)
3.8%	5.4%	Chris Paul (Suns)
N/A	2.1%	Rudy Gobert
1.6%	1.6%	Trae Young
1.5%	1.5%	Jayson Tatum
11.8%	N/A	Ja Morant
2.1%	N/A	DeMar DeRozan

- Basketball-Reference has a secret algorithm
- The sum of the probabilities add up to 100%
- To match that, I used Bayes Theorem
- 31 RF is a probability of candidate given a top ten candidate is selected

What if each algorithm was a voter?

	1st	2nd	3rd	4th	5th	Points
Devin Booker (Suns)	30	0	5	6	0	41
Giannis Antetokounmpo	0	28	10	0	0	38
Ja Morant	20	14	0	0	1	35
Nikola Jokić	10	0	5	3	2	20
Joel Embiid (76ers)	0	0	10	6	1	17
Luka Dončić	0	0	0	3	2	5

1st place vote = 10 points
2nd place vote = 7 points
3rd place vote = 5 points
4th place vote = 3 points
5th place vote = 1 point

- If order doesn't matter
- Top 3 has 1/3 correct
- Top 4 has 3/4 correct
- Top 5 has 4/5 correct
- Top 6 has 5/6 correct

Final Conclusions

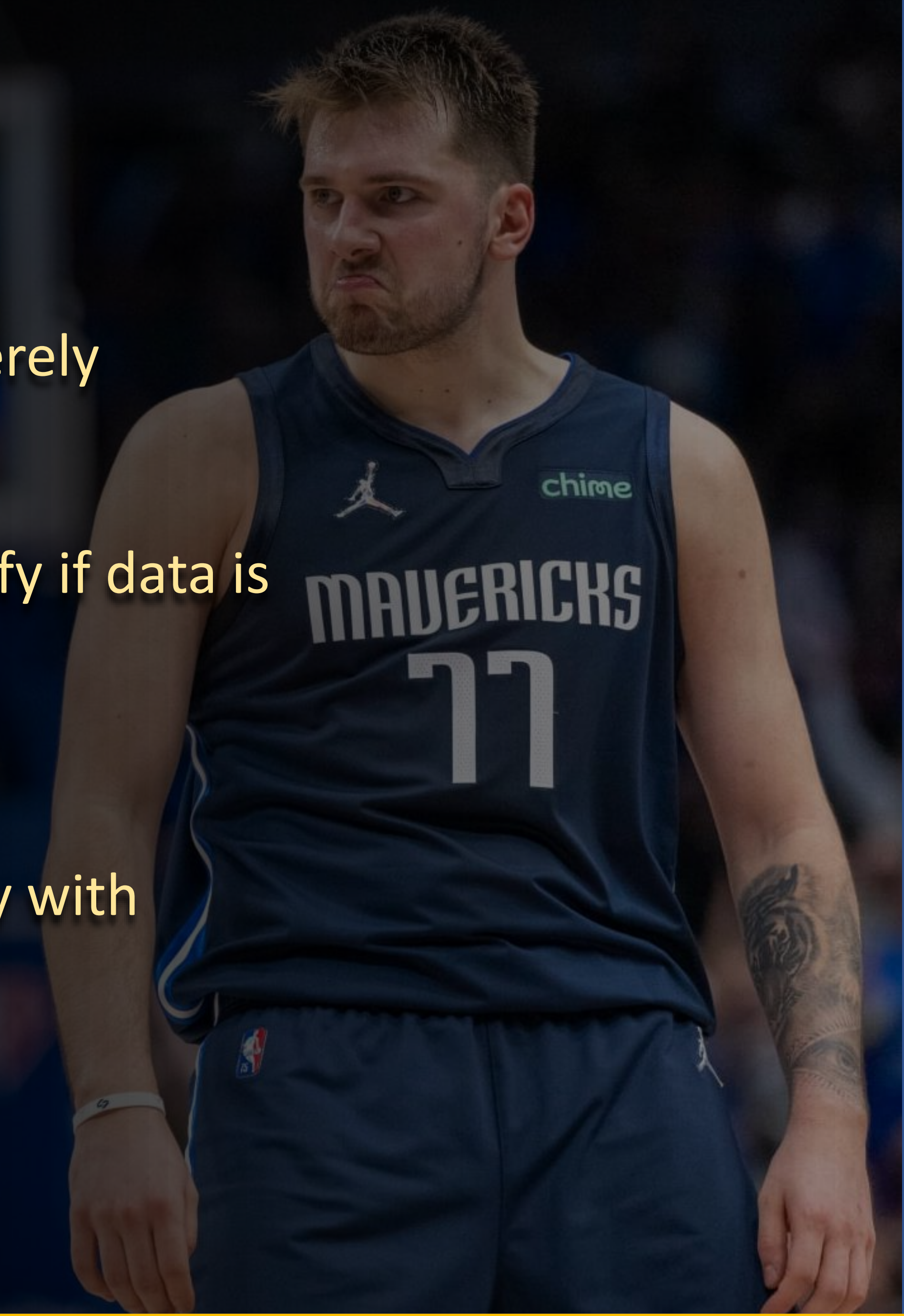
- Performing Principal Component Analysis shows that PC1(~61%) is related to the quality of the player and PC2(~12%) is related to playstyle
- MVP Candidates are further from the origin in a biplot
- Points, free throw attempts, 2 point attempts, Wins are the top 4 stats for a MVP candidate
- Using past 10 years of data may not be effective as using the past 15 years or 31 years
- The importance rank of variables did not change based on classifying techniques: Extra Trees and Random Forest
    - Potentially because we took the average of 100 iterations of the tree

Future Work

- Find alternative ways to trim down 44 variables to 15 variables
- Use 13 variables for PCA instead of 8
- Replace Team winning percentage with player win percentage
  - Affects Ja Morant and James Harden
- Research a predictive model that can pair independent variable with specific response variable such as 1991 stats with 1992 MVP votes
  - Original goal was to predict the MVP one year in advance
- Use decision tree prediction models built into the Scikit Learn package

Future Work

- Undersampling and/or oversampling adjustment for severely imbalanced data
  - # of MVPS is low compared to # of Non-Candidates
  - split data into training/validation/test in order to verify if data is overfitting
- Analyze the predictive performance of our algorithms
  - F1 Score
  - AUC and ROC – better metric over accuracy especially with classification task



References

AGRON Stats. “Biplot using base graphic functions in R.” Agron Info Tech. 26 June 2018, http://agroninfotech.blogspot.com/2020/06/biplot-for-pcs-using-base-graphic.html.Written Accessed 1 May 2022.

Galarnyk, Michael. “Logistic Regression Using Python (Scikit-Learn).” *Medium*, 29 Apr. 2020, towardsdatascience.com/logistic-regression-using-python-sklearn-numpy-mnist-handwriting-recognition-matplotlib-a6b31e2b166a

Raheel Shaikh. “Feature Selection Techniques in Machine Learning with Python.” *Medium*, Towards Data Science, 28 Oct. 2018, towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e

Vinco, Vivo. “1991-2021 NBA Stats.” *Kaggle.* 15 Apr. 2022, www.kaggle.com/datasets/vivovinco/19912021-nba-stats Accessed 1 May 2022.

Yalçın, Orhan G. “Interpretable Machine Learning in 10 Minutes with RuleFit and Scikit Learn.” *Medium*, 24 June 2021, towardsdatascience.com/interpretable-machine-learning-in-10-minutes-with-rulefit-and-scikit-learn-da9ebb925795 Accessed 3 May 2022.

“RuleFit: A Modeling Method for Automatically Extracting Interactions.” *HACARUS INC.*, 5 Jan. 2022, https://hacarus.com/ai-lab/20211208-rulefit/ Accessed 1 May 2022.

Silva, João V.R.d., and Paulo C. Rodrigues. 2022. "All-NBA Teams’ Selection Based on Unsupervised Learning" *Stats* 5, no. 1: 154-171. https://doi.org/10.3390/stats5010011

Thank You!
Questions?