CAPSTONE PROJECT:
NBA PLAYER ANALYSIS & Hall Of Famer Predictions

General Assembly DSI 14
June 29, 2020
Chan Song Yuan

TABLE OF CONTENTS

Problem Statement

For this capstone project, I want to predict which NBA players will have higher probability to get nominated as Hall Of Fame in the near future.

- Which features in the data sets to be chosen ?
- Player performance ?
- Regressions / Classifications?

EXPLORATORY DATA ANALYSIS

EDA helps to identify which features had higher importance for Hall Of Famer Predictions. Some analysis on players performance

Data Sets

- Data sets which obtained from Kaggle
- The kaggle user scraped the data from Basketball Reference.com
- Data available from year 1950 - 2017

Seasons_Stats

- Shape : 24691 x 53
- Players stats recorded in yearly manner
- Glossary as reference for each feature in the data set,

player_data

- Shape : 4550 x 9
- Players positions, height, weight
- Birthdate, college, year start, year end

Players

- Shape : 3921 x 8
- Players height, weight
- Birthdate, college, year born, birth city, birth state

Some Notable Rules Change Due To Player Dominance

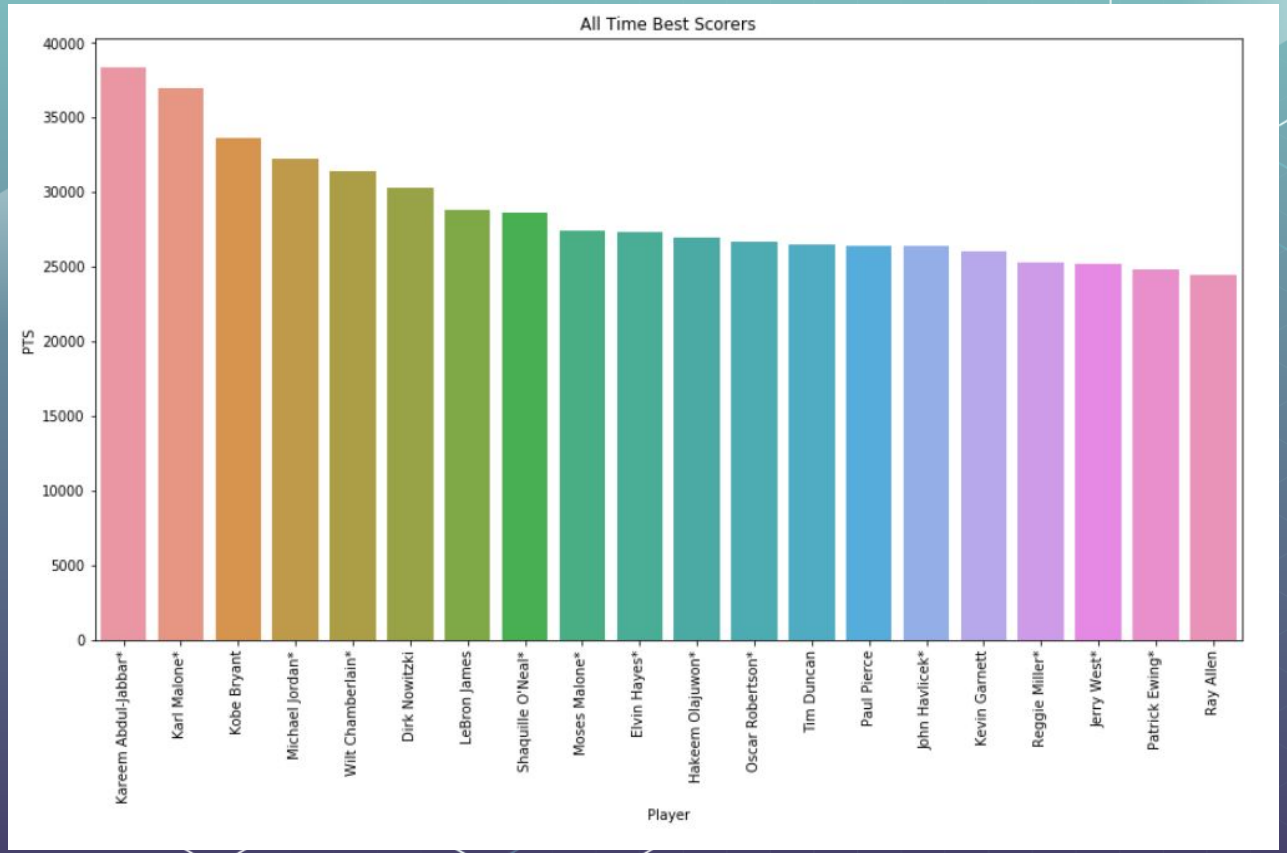
Links

1954-1955	1964-65	2000-2001	2001 - 2002
24 Seconds shot clock and penalty free throw introduced	**Free Throw Line moved from 12 to 16 feet, forbid to cross free throw line when perform FT**	**The Booty Rule / Mark Jackson Rule (5 seconds violations)**	**The Defensive 3-Second Rule**
This is prevent the trailing team to foul deliberately	Wilt would throw it against the backboard for dunks	Offensive player with the ball and not facing-up starts dribbling below the free throw line extended while being closely guarded or starts dribbling outside and then penetrates below the free throw line extended while being closely guarded	Defence player not allowed to stay in the restricted area for more than 3 seconds. (Shaquille O'Neal dominace)

“Just remember Michael. When you played, they changed the rules to make it easier for you to dominate. When I played, they changed the rules to make it harder for me”

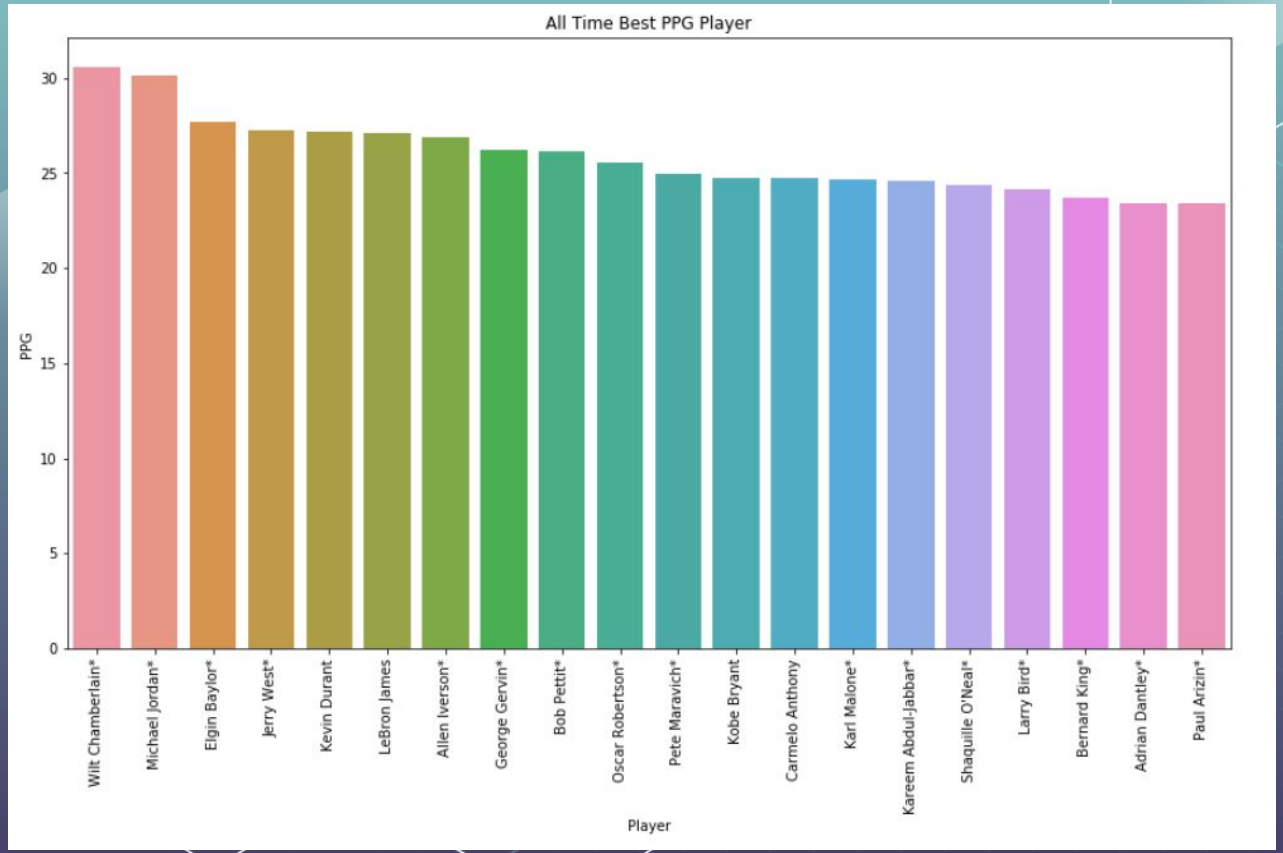
- Wilt Chamberlain

All Time Scorers



	id	Player	PTS
730	788	Kareem Abdul-Jabbar*	38387.0
1591	1736	Karl Malone*	36928.0
2259	2455	Kobe Bryant	33643.0
1536	1674	Michael Jordan*	32292.0
455	493	Wilt Chamberlain*	31419.0
2409	2625	Dirk Nowitzki	30260.0
2689	2943	LeBron James	28787.0
2055	2221	Shaquille O'Neal*	28596.0
1095	1184	Moses Malone*	27409.0
706	761	Elvin Hayes*	27313.0
1540	1678	Hakeem Olajuwon*	26946.0
489	524	Oscar Robertson*	26710.0
2327	2534	Tim Duncan	26496.0
2413	2629	Paul Pierce	26397.0
535	577	John Havlicek*	26395.0
2213	2400	Kevin Garnett	26071.0
1701	1861	Reggie Miller*	25279.0
491	526	Jerry West*	25192.0
1580	1720	Patrick Ewing*	24815.0
52	58	Ray Allen	24505.0

All Time Efficient Scorers



	Player	PPG
185	Wilt Chamberlain*	30.592828
814	Michael Jordan*	30.163303
165	Elgin Baylor*	27.673139
208	Jerry West*	27.237546
1550	Kevin Durant	27.209695
1429	LeBron James	27.121076
1199	Allen Iverson*	26.875923
530	George Gervin*	26.180839
118	Bob Pettit*	26.177270
206	Oscar Robertson*	25.526368
378	Pete Maravich*	24.942625
1191	Kobe Bryant	24.753561
1415	Carmelo Anthony	24.730485
842	Karl Malone*	24.682985
344	Kareem Abdul-Jabbar*	24.587516
1084	Shaquille O'Neal*	24.343172
636	Larry Bird*	24.123336
593	Bernard King*	23.686373
518	Adrian Dantley*	23.422634
58	Paul Arizin*	23.402477

Scoring Efficiency vs Total Points

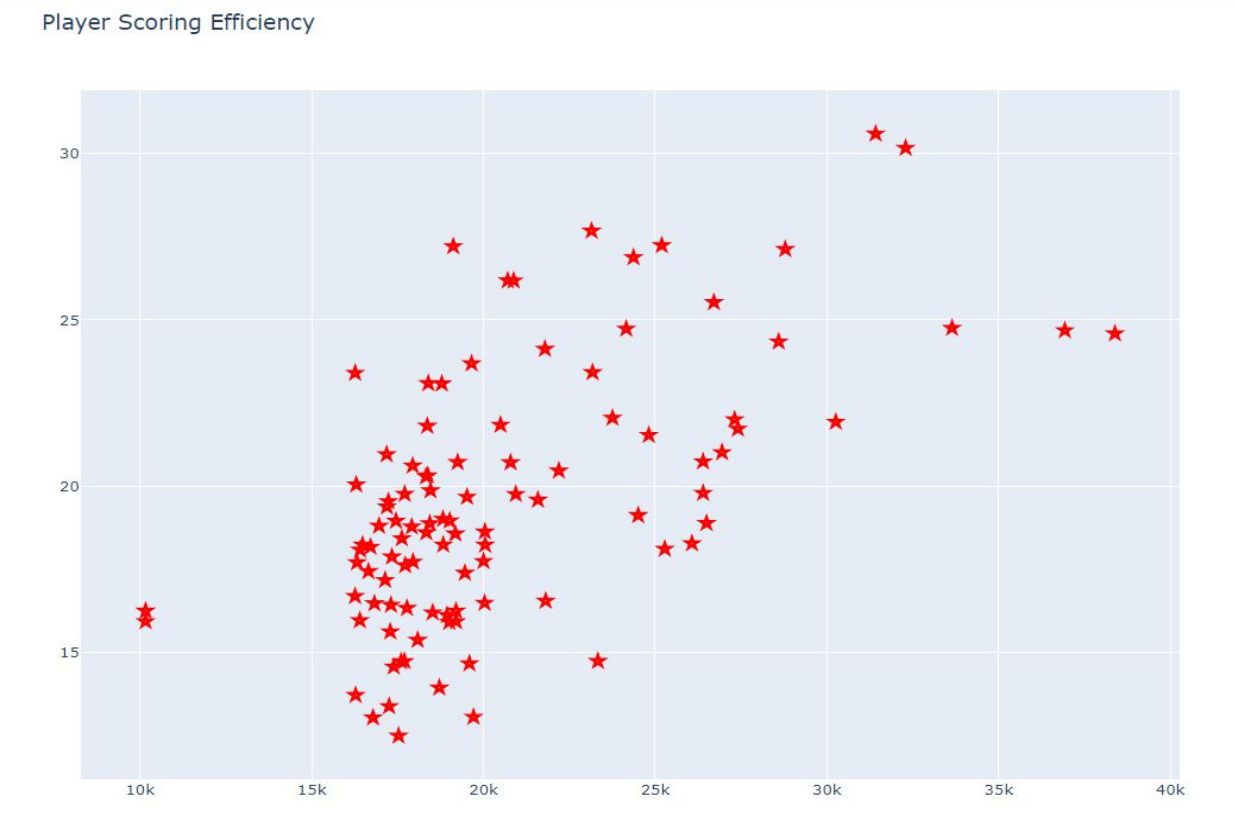
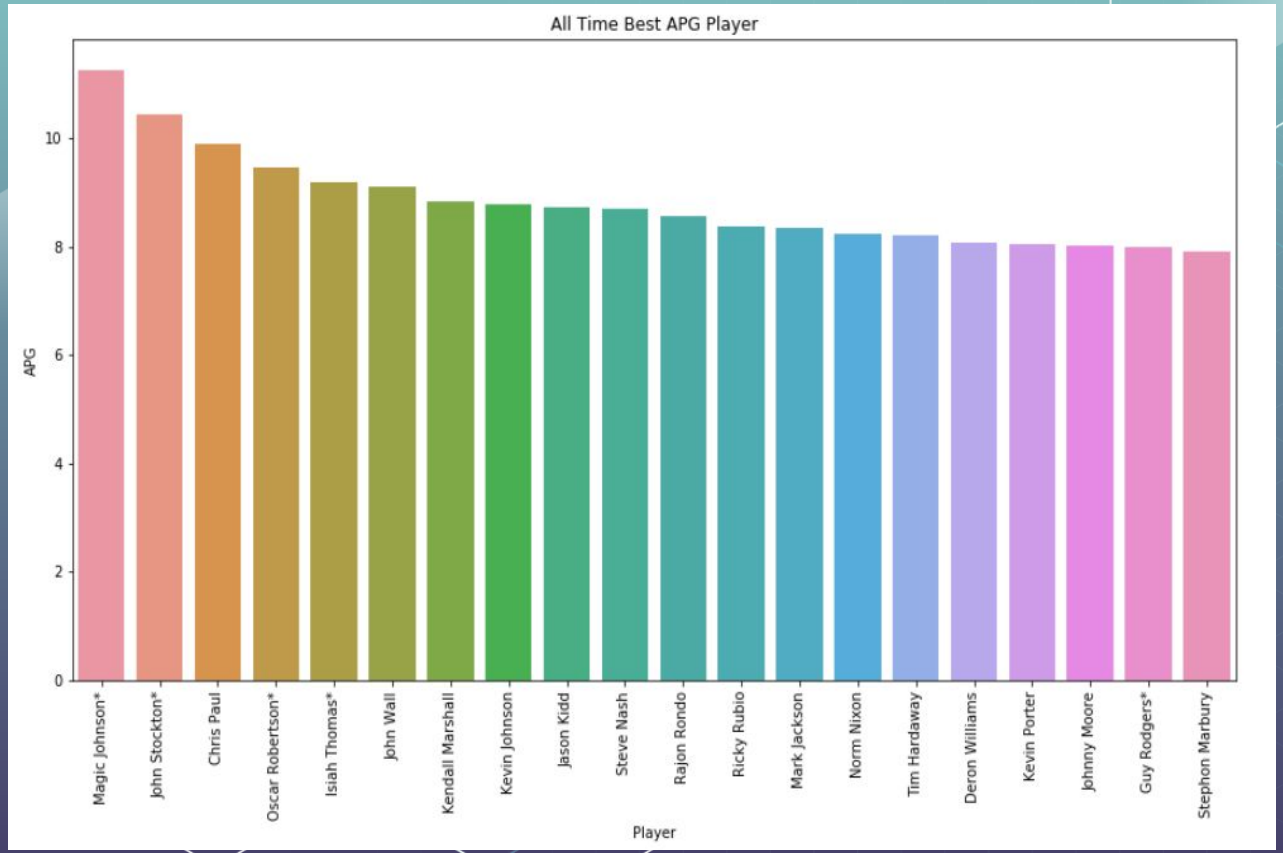


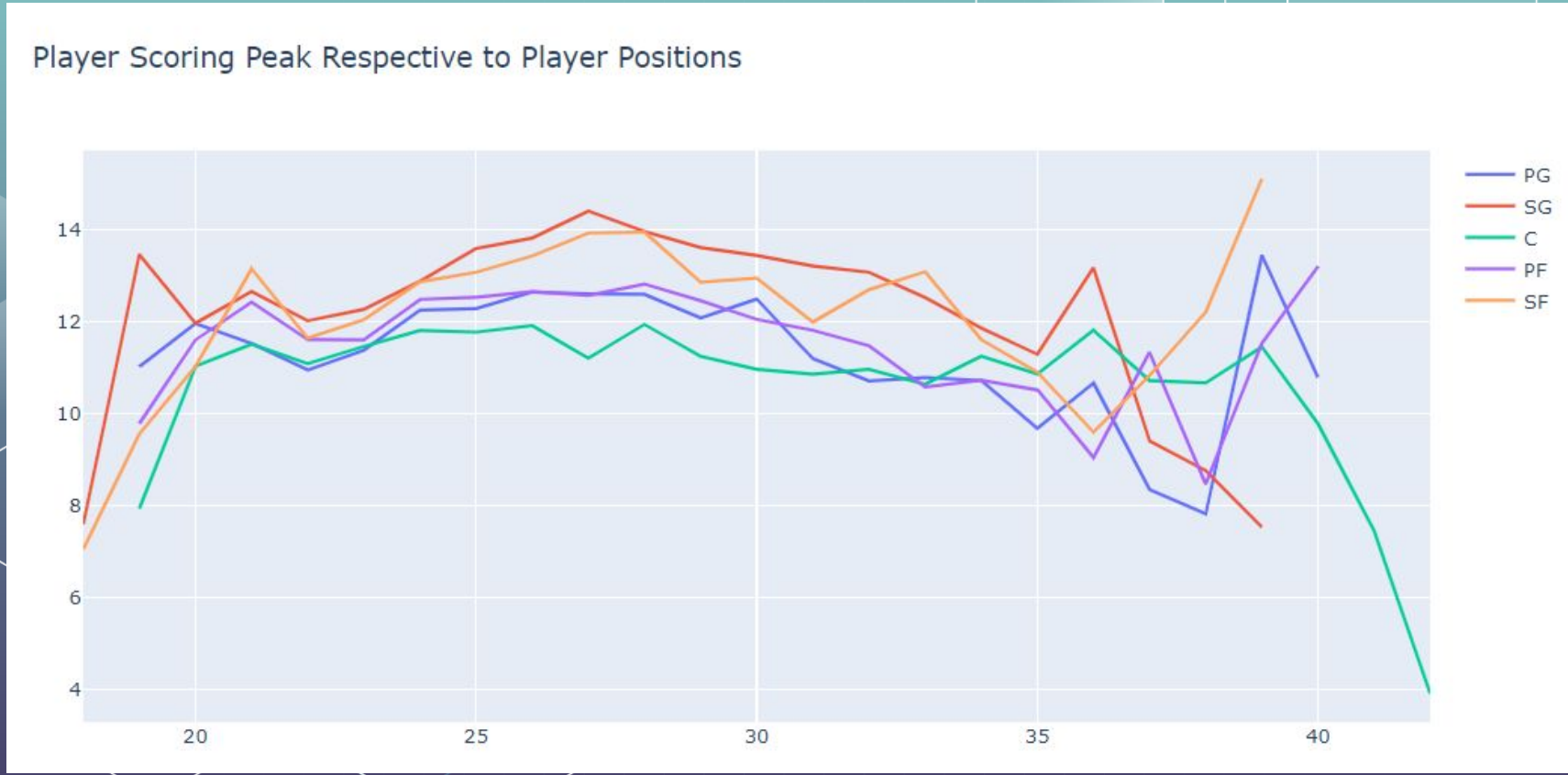
Tableau Link

All Time Efficient Assistant



	Player	APG
652	Magic Johnson*	11.257271
822	John Stockton*	10.434039
1499	Chris Paul	9.884820
206	Oscar Robertson*	9.451219
731	Isiah Thomas*	9.188279
1664	John Wall	9.097141
1713	Kendall Marshall	8.833333
982	Kevin Johnson	8.791257
1133	Jason Kidd	8.724820
1207	Steve Nash	8.687996
1537	Rajon Rondo	8.575619
1684	Ricky Rubio	8.373010
899	Mark Jackson	8.356182
599	Norm Nixon	8.248750
962	Tim Hardaway	8.215419
1507	Deron Williams	8.086645
432	Kevin Porter	8.034819
687	Johnny Moore	8.010578
181	Guy Rodgers*	7.979648
1204	Stephon Marbury	7.898547

Player Scoring Peak Age On Different Position

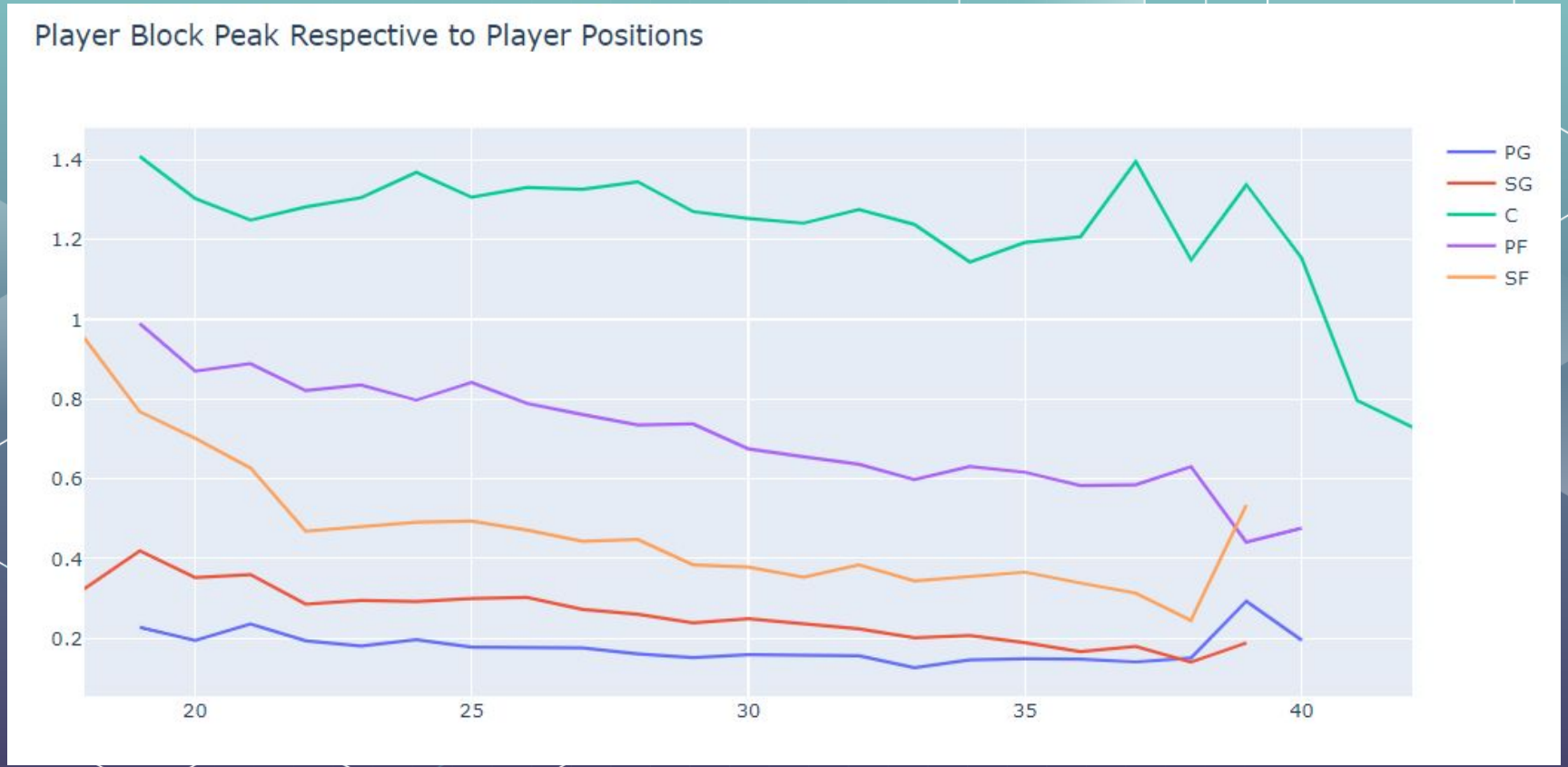


Player Assist Peak Age On Different Position

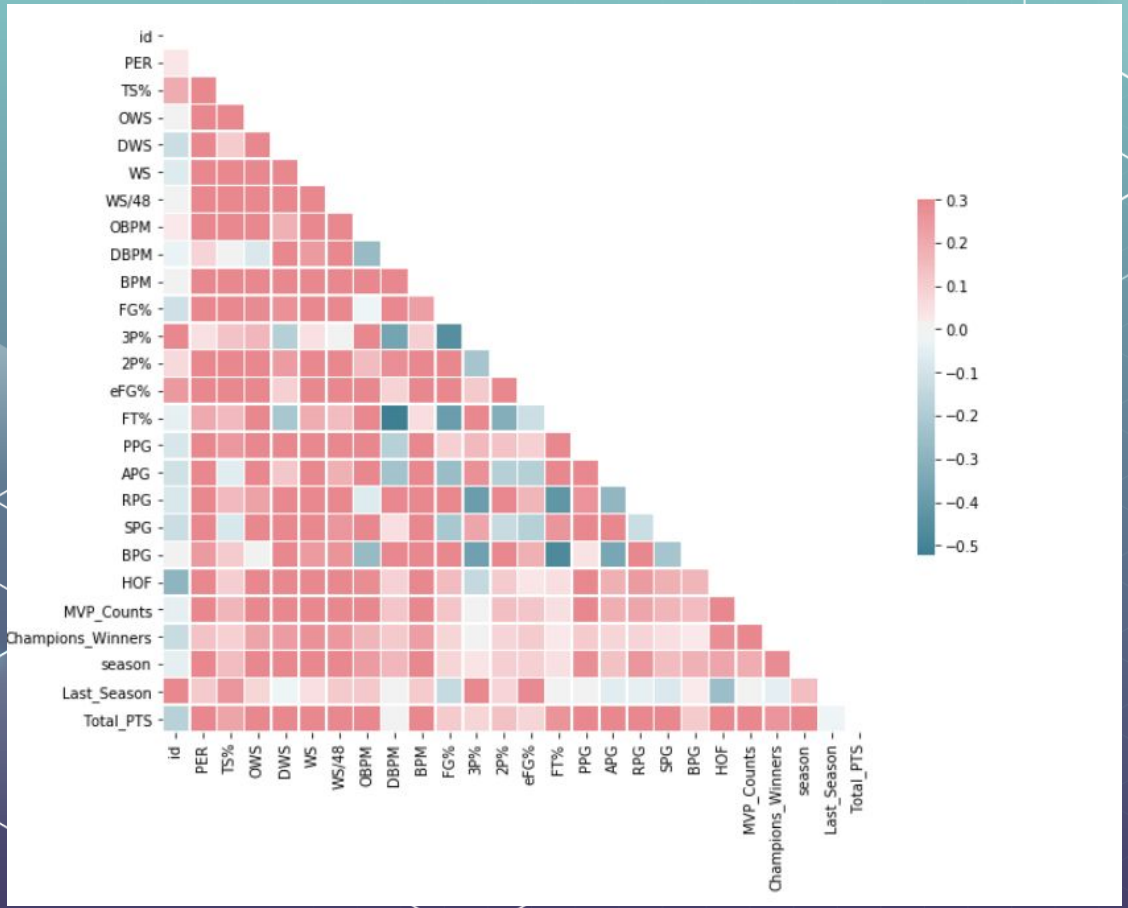
Player Rebound Peak Age On Different Position

Player Steal Peak Age On Different Position

Player Blocking Peak Age On Different Position



Correlations Heatmap



HOF	1.000000
Total_PTS	0.490353
WS	0.427507
PPG	0.395153
OWS	0.371861
PER	0.366142
DWS	0.335137
WS/48	0.332253
BPM	0.319985
MVP_Counts	0.317305
OBPM	0.283787
Champions_Winners	0.279153
RPG	0.247877
season	0.210945
SPG	0.183753
APG	0.181269
BPG	0.167144
FG%	0.150412
2P%	0.105791
TS%	0.102672
DBPM	0.089663
FT%	0.054924
eFG%	0.033950
3P%	-0.140555
Last_Season	-0.254583
id	-0.290965

Name: HOF, dtype: float64

MODELING AND RESULT

Modeling Analysis

Random Forest Classifier

Train Score : 0.9105

id	Player	HOF
2943	LeBron James	0.829534
2625	Dirk Nowitzki	0.818191
3219	Kevin Durant	0.807564
2455	Kobe Bryant	0.796548
2534	Tim Duncan	0.794496
2400	Kevin Garnett	0.789097
3095	Chris Paul	0.752076
2629	Paul Pierce	0.751899
2554	Tracy McGrady	0.749821
3324	Russell Westbrook	0.739049
58	Ray Allen	0.737332
2807	Pau Gasol	0.725139
2906	Carmelo Anthony	0.689502
2999	Dwight Howard	0.676795
2490	Steve Nash	0.673197
2915	Chris Bosh	0.672489
2485	Stephon Marbury	0.660598
3342	Stephen Curry	0.656834
3342	Willie Reed	0.656834
2522	Chauncey Billups	0.653814

AdaBoost Classifier

Train Score : 1.0

id	Player	HOF
2534	Tim Duncan	0.620322
2400	Kevin Garnett	0.602696
2943	LeBron James	0.600571
3219	Kevin Durant	0.589252
3342	Stephen Curry	0.581682
3342	Willie Reed	0.581682
2455	Kobe Bryant	0.572167
2999	Dwight Howard	0.568905
2625	Dirk Nowitzki	0.568744
3095	Chris Paul	0.558395
3324	Russell Westbrook	0.558192
2906	Carmelo Anthony	0.545597
2490	Steve Nash	0.541841
2522	Chauncey Billups	0.538627
58	Ray Allen	0.537039
2485	Stephon Marbury	0.536454
2822	Tony Parker	0.535199
2336	Jason Kidd	0.534372
2859	Manu Ginobili	0.533430
2807	Pau Gasol	0.532521

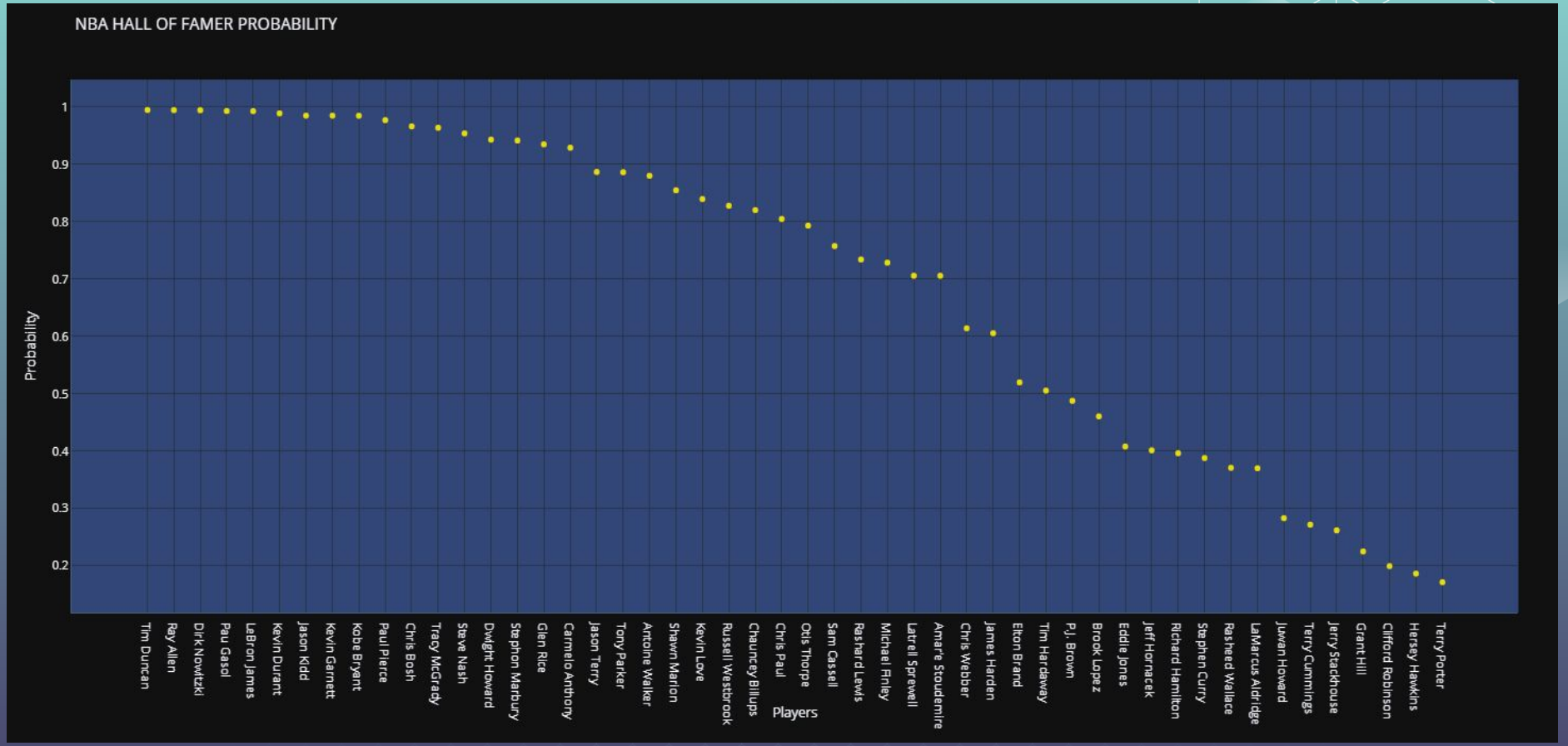
XGBoost Classifier

Train Score : 1.0

id	Player	HOF
2534	Tim Duncan	0.994822
58	Ray Allen	0.994599
2625	Dirk Nowitzki	0.994424
2807	Pau Gasol	0.992963
2943	LeBron James	0.992784
3219	Kevin Durant	0.988898
2336	Jason Kidd	0.984888
2400	Kevin Garnett	0.984877
2455	Kobe Bryant	0.984721
2629	Paul Pierce	0.977102
2915	Chris Bosh	0.966249
2554	Tracy McGrady	0.963825
2490	Steve Nash	0.953981
2999	Dwight Howard	0.943091
2485	Stephon Marbury	0.941559
2027	Glen Rice	0.934909
2906	Carmelo Anthony	0.929067
2699	Jason Terry	0.886771
2822	Tony Parker	0.886262
4180	Antoine Walker	0.880008

Modeling Analysis

MODELS	BEST PARAMS	Train Score	Results Analysis
Random Forest	'rf__max_depth': 2, 'rf__max_leaf_nodes': 3, 'rf__min_samples_leaf': 1, 'rf__n_estimators': 10000	0.9105	The train score is lower but the results seems acceptable, but by comparing with XGBoost Classifier model, the probability results is lower.
AdaBoost	'ada__learning_rate': 0.7, 'ada__n_estimators': 10000	1.0	Even though the train score is 1.0, but the probability were lower than 0.7
XGBoost	eval_metric = ‘auc’, scale_pos_weight = 0.4, 'xg__learning_rate': 0.6	1.0	The train score is 1.0 and comparing the results, the probability by XGBoost Classifications seems perform better than other 2 models



https://plotly.com/~songyuan89/8/

CONCLUSIONS

THANKS