

NBA Players Examination and Salaries Prediction

Calvin Yu

Motivation



- Interested in studying players' performance throughout the years
- Interested in examining how players' performance influence to the salaries

Impact Hypothesis



- Create a machine learning model to predict players' salaries based on players' stats
- Team GMs and NBA fans can know how much a player ought to get based on their performance
- Create interactive dashboards to visualize Players stats for people who are interested in knowing

Solution Path

1. Web Scraped and downloaded all the data that will be used
2. Data Wrangling : Fill up the missing values, convert the data type, etc.,
3. Create a function that takes different algorithms and select the most accurate one
4. Return the result of the function above and store it as a dataframe, and visualize it in tableau
5. Create interactive dashboards to present

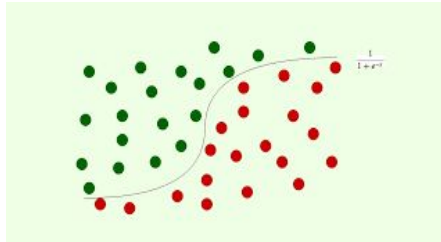


Data

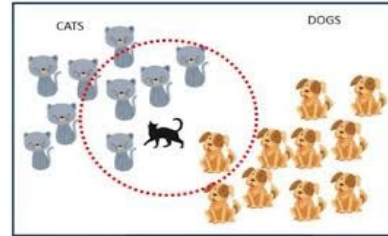


- Web Scraped from [basketball-reference.com](https://www.basketball-reference.com) and downloaded from [Kaggle.com](https://www.kaggle.com)
- 6225 rows with 58 columns
- Feature highlight : Salary, Predicted Salary, eFG%, 3P%, Ast%, Reb%, Pts
- Take all the data points and set cross fold to 10 , so all the data points can be trained and used for prediction

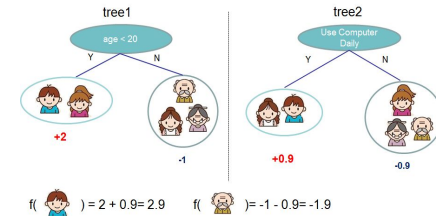
Algorithms



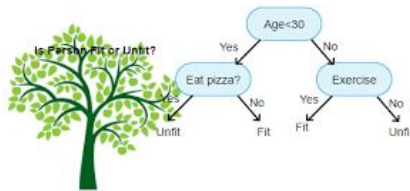
Logistic
Regression



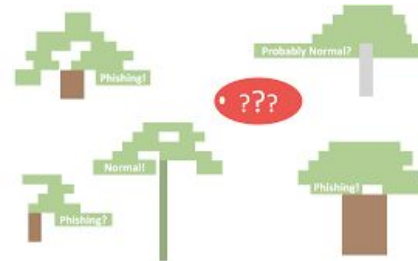
KNN



Gradient
Boost /
XGBoost



Decision
Tree



Random
Forest

Results

Linear Regression: mean R2 score = 0.49 (std = 0.29)

KNN: mean R2 score = 0.47 (std = 0.21)

Decision Tree: mean R2 score = 0.27 (std = 0.20)

Random Forest: mean R2 score = 0.64 (std = 0.10)

Gradient Boosting: mean R2 score = 0.63 (std = 0.16)

	Player	Ht	Wt	Age	Pos_pref	Colleges	Year_Play	Tm	G	GS	...	WS/48	OBPM	DBPM	BPM	VORP	Year	salary	Predicted_Salar
0	Arron Afflalo	1.9558	210.0	22	SG	UCLA	1	DET	75	9	...	0.092	-2.6	1.0	-1.5	0.1	2008	1015440	1170556.8
1	Arron Afflalo	1.9558	210.0	23	SG	UCLA	2	DET	74	8	...	0.069	-2.8	0.2	-2.6	-0.2	2009	1086240	1306202.2
2	Arron Afflalo	1.9558	210.0	24	SG	UCLA	3	DEN	82	75	...	0.092	-0.2	-0.2	-0.4	0.9	2010	1959577	3754399.6
3	Arron Afflalo	1.9558	210.0	25	SG	UCLA	4	DEN	69	69	...	0.128	1.7	-0.3	1.4	2.0	2011	7562500	7208737.5
4	Arron Afflalo	1.9558	210.0	26	SG	UCLA	5	DEN	62	62	...	0.121	1.4	-1.2	0.1	1.1	2012	7562500	7628156.2
...
6220	Devin Vassell	1.9558	200.0	20	SF	Florida State	1	SAS	62	7	...	0.069	-2.2	0.7	-1.5	0.1	2021	4235160	3665451.6
6221	Patrick Williams	2.0066	215.0	19	PF	Florida State	1	CHI	71	71	...	0.060	-2.8	0.4	-2.4	-0.2	2021	7422000	5590864.7
6222	Dylan Windler	1.9812	196.0	24	SF	Belmont	1	CLE	31	0	...	0.045	-2.0	0.3	-1.7	0.0	2021	2239200	2130593.5
6223	Cassius Winston	1.8542	185.0	22	PG	Michigan State	1	WAS	22	0	...	0.066	-2.1	-1.1	-3.2	0.0	2021	462629	556877.1



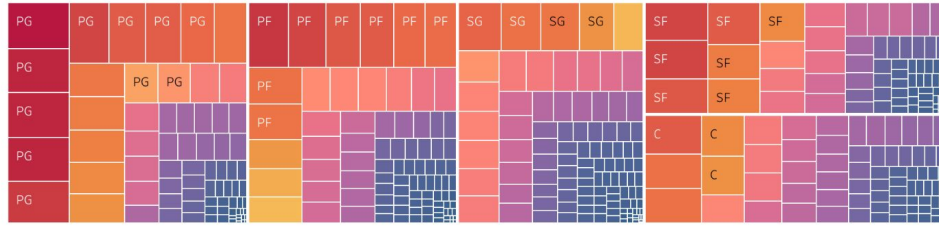
Best Model is Random Forest

- Can explain 64% of the variance

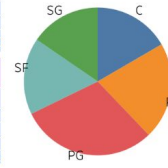
(Not bad considering contracts are not signed based on current season but the season before)

Interactive Dashboards

NBA Player Salary



Average Salary by Position per Year



Year
2021 to 2021
and Null values

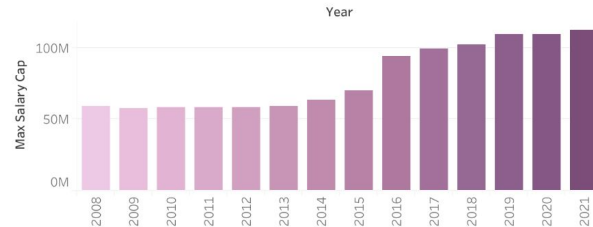
Position

C
PF
PG
SF
SG

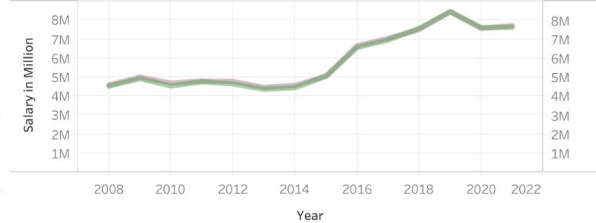
Avg. Salary

19,186 45,780,966

NBA salary cap



Actual Salary vs Prediction Salary



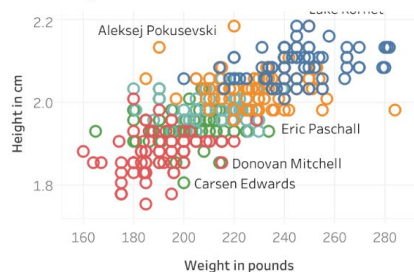
Year
2008 2021

Measure Names

Avg. Predicted Salary
Avg. Salary

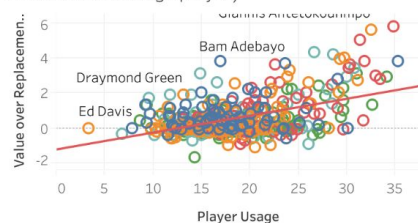
Interactive Dashboards con.

NBA Player Size

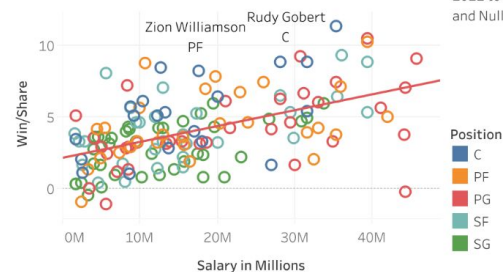


Value over Replacement Player

(How much better when a team substitute him in instead of a average player)



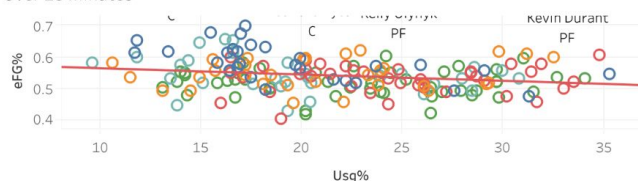
Studs vs Duds



Year
2021 to 2021
and Null values

Most Effective Scorer

Scoring Efficiency (eFG%) vs Total Usage of Play (Usg%)
Over 25 Minutes



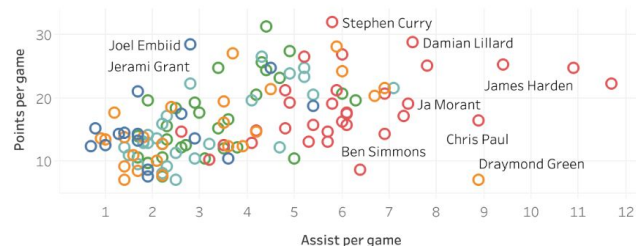
Best Defensive Player

Block Rate (an estimate of the percentage of opponent possessions that end with a block by the player while he was on the floor) VS Steal Rate (Steal Percentage is an estimate of the percentage of opponent possessions that end with a steal by the player while he was on the floor)

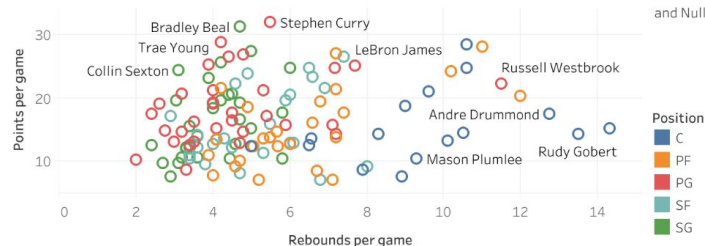


Interactive Dashboards con.

Points / Assist (Over 25mins per game)

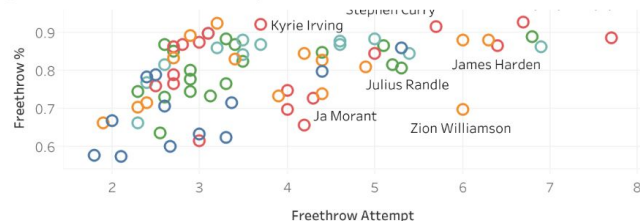


Points / Rebounds (Over 25mins per game)

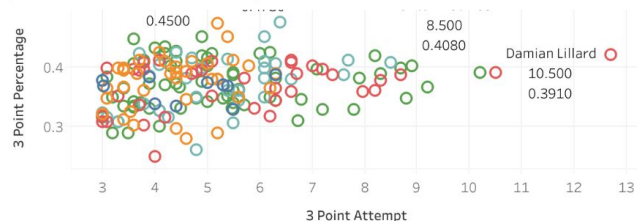


Year
2021 to 2021
and Null values

Best Freethrow Shooters
(At least 3 attempts per game)



Best 3 Points Shooters
(At least 3 attempts per game)



Communication



- NBA contracts are not signed based on the current season's performance but signed year or years before, so can't avoid players who fall off after signing a big contract (JOHN WALL, CHANDLER PARSONS)
- The interactive dashboards can examine different fields of the dataset
- The averaged salary of each year and the predicted average salary of each year are not differ much



Further step

- Try to gather the marketing dataset from NBA
- Believe that players' products sales is an important piece in predicting their salary
- Players' business value is important too (Lebron James, Kobe Bryant, Michael Jordan)
- Merge the marketing dataset into the dataset that I have to see if there is any improvement on the machine learning model