

MoneyBBall

Predicting the market value of NBA players

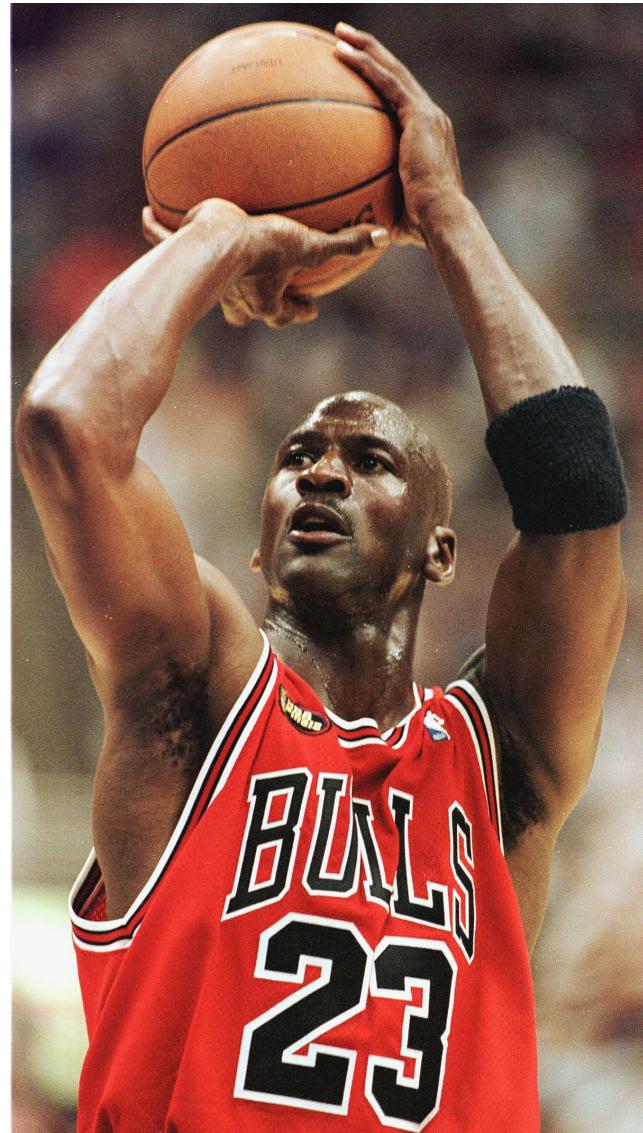


Metis Data Engineering Module Presentation
Andrei Levin • 12/17/21

How much is an NBA player worth?

Salary

- Players sign contracts for 1-5 years, between ~\$1-40M/year
- A player may over- or under-perform his contract (especially a few years in)



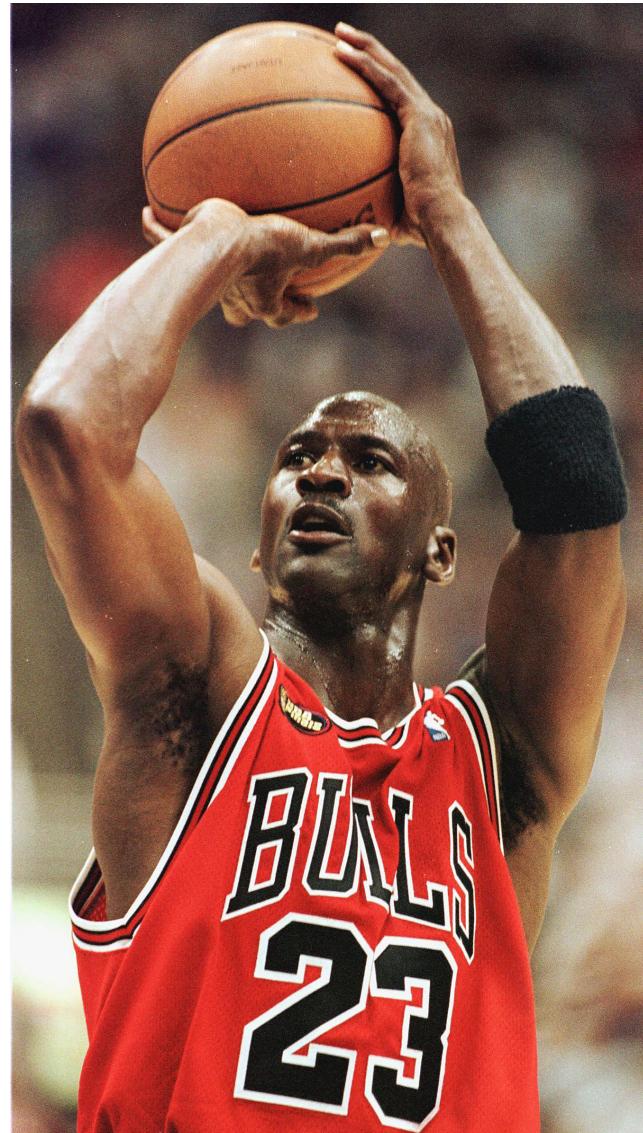
How much is an NBA player worth?

Salary

- Players sign contracts for 1-5 years, between ~\$1-40M/year
- A player may over- or under-perform his contract (especially a few years in)

Market Value

- If this player was a free agent today, what size contract would they get?
- Potentially useful metric!



Sounds like a good problem for Machine Learning!

The Big Idea: track all free agents for a given season

- **Target:** What was his salary the following year?
(data: scrape *hoopshype.com*)
- **Features:** What were his stats the previous year?
(data: scrape *basketball-reference.com*)

Sounds like a good problem for Machine Learning!

The Big Idea: track all free agents for a given season

- **Target:** What was his salary the following year?
(data: scrape *hoopshype.com*)
- **Features:** What were his stats the previous year?
(data: scrape *basketball-reference.com*)
- Do this for each free agent for the past 5 years (~500 records)
- Train a regression model

BASKETBALL REFERENCE

Enter Person, Team, Section, etc

Players Teams Seasons Leaders Scores Playoffs Draft Stathead Newsletter Full Site Menu Below ▾

2019 Free Agent Tracker

287 Free Agents Share & Export ▾ Glossary

Rk	Player	Pos	Age	Type	OTm	2018-19 Stats	WS	NTm	Terms	Notes
1	Kevin Durant	F-G	32-283	UFA	GSW	26.0 Pts, 6.4 Reb, 5.9 Ast	11.5	BRK	Signed 4-yr \$164.25M contract with GSW as part of sign-and-trade	
2	Nikola Vučević	C	30-258	UFA	ORL	20.8 Pts, 12.0 Reb, 3.8 Ast	10.1	ORL	Signed 4-yr \$100M contract	
3	Kawhi Leonard	F	30-010	UFA	TOR	26.6 Pts, 7.3 Reb, 3.3 Ast	9.5	LAC	Signed 3-yr \$103.13M contract	
4	Kyrie Irving	G	29-108	UFA	BOS	23.8 Pts, 5.0 Reb, 6.9 Ast	9.1	BRK	Signed 4-yr \$136.49M contract	
5	Jimmy Butler	F-G	31-298	UFA	PHI	18.7 Pts, 5.3 Reb, 4.0 Ast	7.9	MIA	Signed 4-yr \$140.79M contract with PHI as part of sign-and-trade	
6	Al Horford	C-F	35-036	UFA	BOS	13.6 Pts, 6.7 Reb, 4.2 Ast	7.5	PHI	Signed 4-yr \$109M contract	
7	Kemba Walker	G	31-062	UFA	CHO	25.6 Pts, 4.4 Reb, 5.9 Ast	7.4	BOS	Signed 4-yr \$140.79M contract with CHO as part of sign-and-trade	

Features

BASKETBALL REFERENCE

Enter Person, Team, Section, etc

Kevin Durant

Kevin Wayne Durant • Twitter: [KDTray5](#)
(Durant, KD, Slim Repper, Easy Money Sniper, The Servant, Green Room)
Positions: Small Forward, Power Forward, and Shooting Guard • Shoots: Right
6-10, 240lb (206cm, 108kg)
Born: September 29, 1988 (Age: 32-283d) in Washington, District of Columbia

College: [Texas](#)
High School: Monroe Christian School in Rockville, Maryland
Recruiting Rank: 2006 (8)
Draft: [2007 NBA Draft](#), 1st round (2nd pick, 2nd overall), 2007 NBA Draft
NBA Debut: October 31, 2007
Experience: 12 years

Target

→HOOPSHYPE

2019/20 NBA Player Salaries

GO IN-DEPTH

- Salaries of players right now
- Salaries of players back in the day
- Ranking players by wages
- Upcoming free agents
- All NBA awards in history
- NBA DK ratings
- Re-ordering draft classes
- TV ratings
- Listen to our podcast
- Follow us on Instagram

TEAM PAYROLLS PLAYER SALARIES

2019/20

PLAYER	2019/20	2019/2020
1. Stephen Curry	\$40,231,758	\$40,491,547
2. Chris Paul	\$38,500,482	\$38,755,151
3. Russell Westbrook	\$38,500,482	\$38,755,151
4. John Wall	\$38,199,000	\$38,441,953

Training a Linear Regression Model

Data Cleaning

- Processing: heights, salaries, names
- Missing data: dropping/imputation
- Important: normalize salaries to salary cap!



Training a Linear Regression Model

Data Cleaning

- Processing: heights, salaries, names
- Missing data: dropping/imputation
- Important: normalize salaries to salary cap!

Feature Engineering

- Ratios:
 $(\text{Games Started} / (\text{Games Played}))$
 $(\text{Games Played}) / (\text{Total Team Games})$
- Polynomials:
 $(\text{Points per game})^2$
 $(\text{Minutes per game})^2$



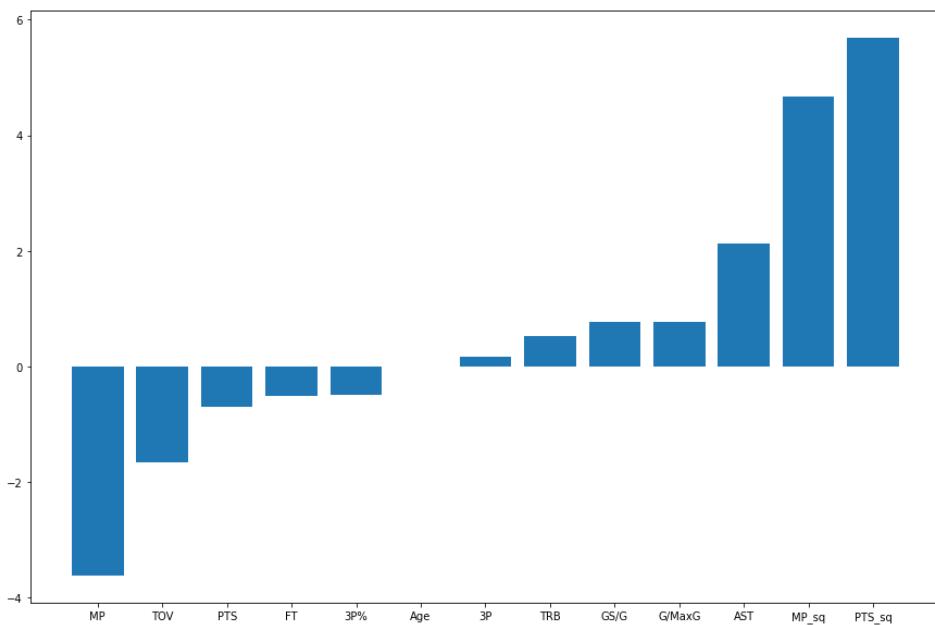
Cross-Validated LASSO Model

LASSO R2 Train 0.7055328969457955

LASSO R2 Test 0.570911491412357

LASSO MAE Test 3.0682963173739832 (\$M)

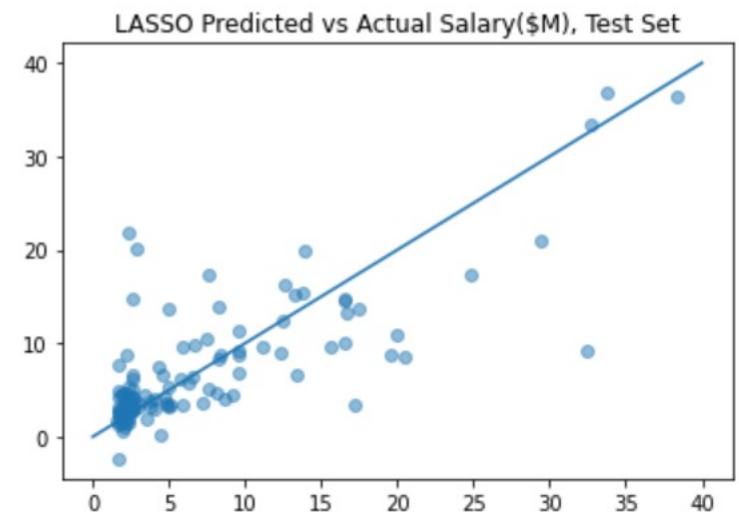
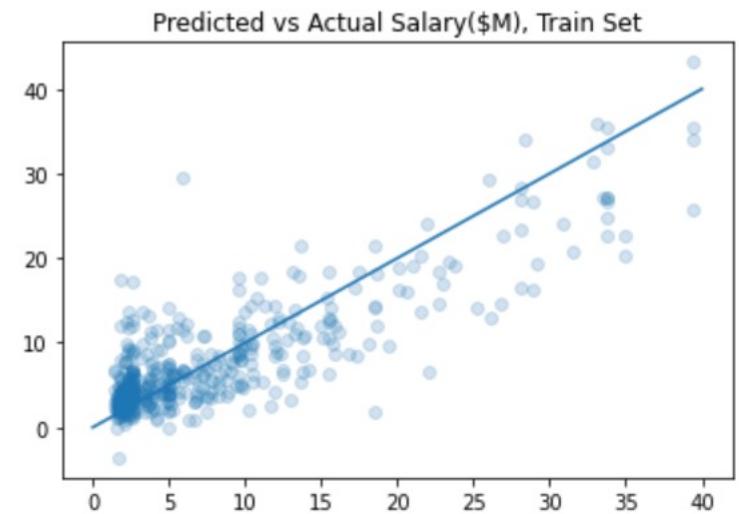
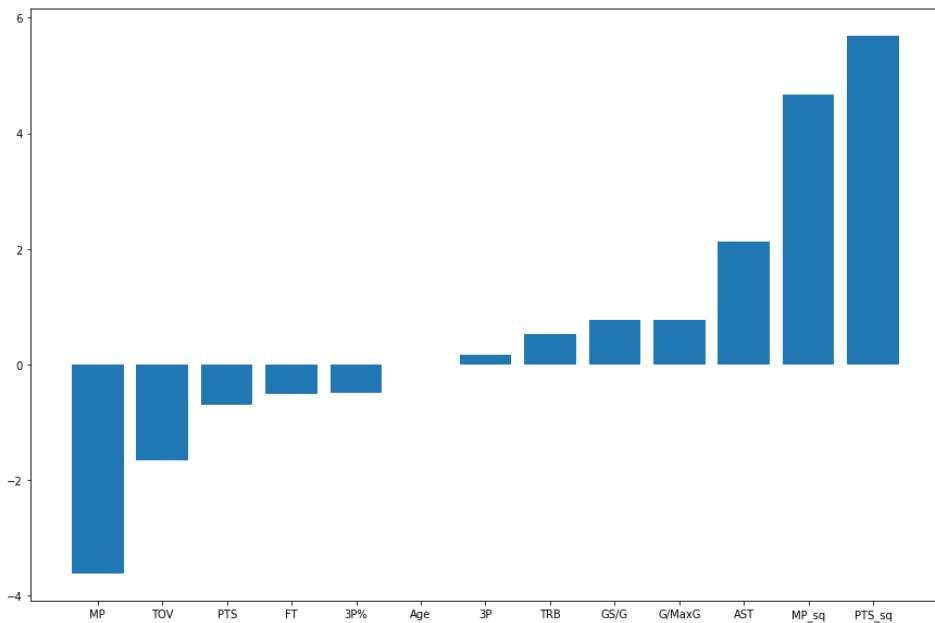
LASSO RMSE Test 5.007879158437053 (\$M)



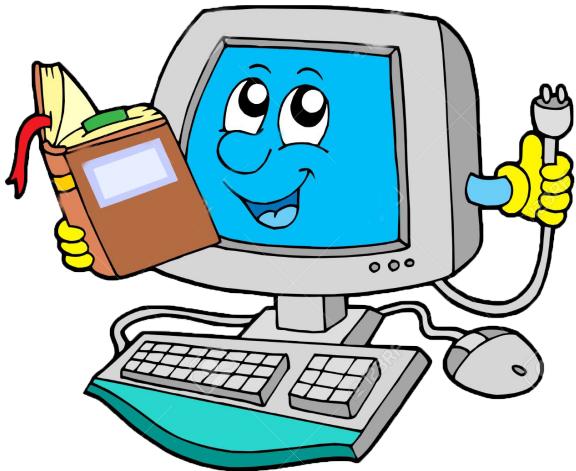
Cross-Validated LASSO Model

LASSO R2 Train 0.7055328969457955
LASSO R2 Test 0.570911491412357

LASSO MAE Test 3.0682963173739832 (\$M)
LASSO RMSE Test 5.007879158437053 (\$M)



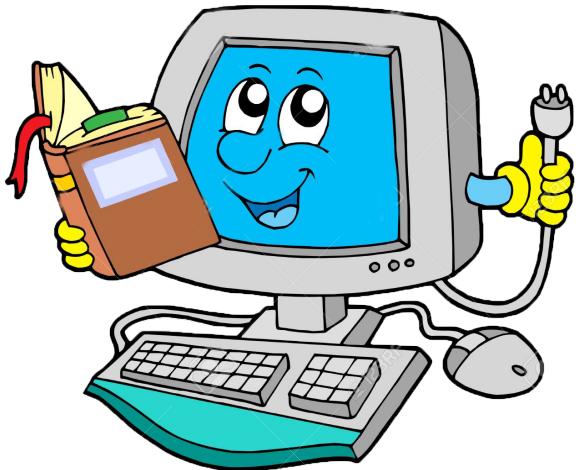
Web App Workflow



Once a Year

- Scrape previous years stats and salaries, save data
- Train regression model, pickle model parameters

Web App Workflow



Once a Year

- Scrape previous years stats and salaries, save data
- Train regression model, pickle model parameters

Once a Day

- Update current season stats for all players
- Predict current market values using pickled model parameters
- Push to github (auto deployed on streamlit.io)



Streamlit

MoneyBBall App Demo

The screenshot shows a Mac OS X desktop with a Chrome browser window open to the MoneyBBall app. The browser's address bar displays the URL: https://share.streamlit.io/andreilevin/dataeng_project/main/streamlit-app.py. The title bar of the browser window says "Streamlit". The Streamlit app interface includes a sidebar on the left with dropdown menus for "Team:" (set to "Atlanta Hawks") and "Player:" (set to "Bogdan Bogdanovic"). The main content area features a title "MoneyBBall" with a subtitle "Predict the market value of NBA players from in-season stats". It displays a message: "This year, Bogdan Bogdanovic is earning a salary of \$18.0M" and "According to our model, his market value is \$12.3M". Below this, a section titled "What about his teammates?" shows a table for the Atlanta Hawks. The table has columns for Tm, Age, Pos, Salary (\$M), Value (\$M), G, GS, MPG, FG%, PTS, AST, and TRB. The data for the team is as follows:

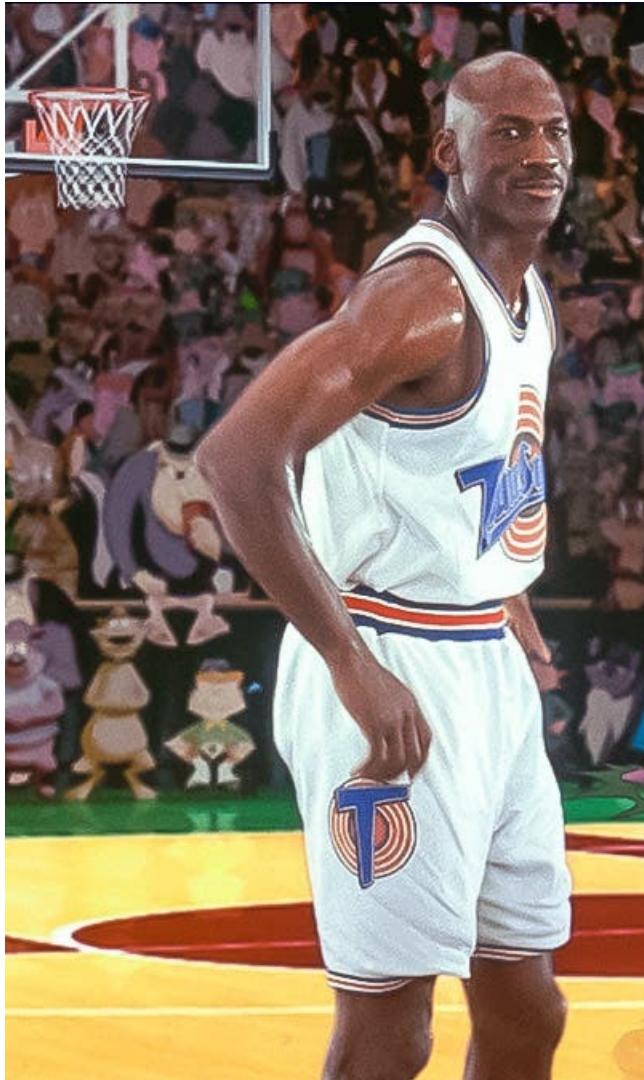
	Tm	Age	Pos	Salary (\$M)	Value (\$M)	G	GS	MPG	FG%	PTS	AST	TRB
Trae Young	ATL	23	PG	8.3	39.7	28	28	34.2	0.461	27.0	9.3	4.1
John Collins	ATL	24	PF	23.0	20.4	28	28	32.4	0.557	17.1	2.1	7.8
Clint Capela	ATL	27	C	17.1	16.5	28	28	30.0	0.570	11.4	1.4	12.9
Bogdan Bogdanovic	ATL	29	SG	18.0	12.3	20	20	28.2	0.447	11.6	2.5	3.6
Kevin Huerter	ATL	23	SG	4.3	10.3	27	17	27.4	0.467	10.7	2.6	3.3
Daniilo Gallinari	ATL	33	PF	20.5	6.3	25	0	21.8	0.425	9.6	1.4	4.0
De'Andre Hunter	ATL	24	SF	7.8	6.1	11	11	27.8	0.450	10.8	0.6	2.7
Sharife Cooper	ATL	20	PG	nan	5.9	6	0	2.3	0.333	0.7	0.7	0.0
Delon Wright	ATL	29	SG	8.5	4.9	26	0	14.3	0.368	2.8	2.2	2.7
Cam Reddish	ATL	22	SF	4.7	4.6	24	0	22.6	0.406	11.3	0.9	

A "Manage app" button is visible at the bottom right of the Streamlit app window. The Mac OS X dock at the bottom contains icons for various applications like Finder, Safari, Mail, and others.

Try it yourself at:

https://share.streamlit.io/andreilevin/dataeng_project/main/streamlit-app.py

Further Work



Modeling

- Advanced stats for better accuracy
- Moving beyond Linear Regression
(Random Forests, XGBoost)

Web App

- Show similarly valued players
- Filters: Age, Position, etc.
- Surplus value calculations
- *What do you want to see?*