Show Me the Money

Exploring and Predicting MLB Free Agent Contracts

Brian Newborn - April 2018

The Question

The Client

How are we going to attack this question?

With Data Science!

Tools and Resources

| | Dashboard | Standard | Advanced | Batted Ball | Win Probability | Pitch Type | Pitch Value | Plate Discipline | Value | Pitch Info |

Show Filters | Custom Reports Export Data

|◄ ◄ 1 2 3 4 5 6 7 8 9 10 ... ► ►| Page size: 30 ▼ 435 items in 15 pages

#	Name	Team	PA	BB%	K%	BB/K	AVG	OBP	SLG	OPS	ISO	Spd	BABIP	UBR	wGDP	wSB	wRC	wRAA	wOBA	wRC+
1	Mike Trout	Angels	507	18.5 %	17.8 %	1.04	.306	.442	.629	1.071	.323	6.2	.318	1.0	0.7	2.3	111	49.6	.437	181
2	J.D. Martinez	- - -	489	10.8 %	26.2 %	0.41	.303	.376	.690	1.066	.387	4.7	.327	-3.4	-2.9	0.5	105	45.1	.430	166
3	Aaron Judge	Yankees	678	18.7 %	30.7 %	0.61	.284	.422	.627	1.049	.343	4.8	.357	0.5	-0.1	-0.4	145	62.5	.430	173
4	Joey Votto	Reds	707	19.0 %	11.7 %	1.61	.320	.454	.578	1.032	.258	2.8	.321	-10.1	0.5	0.0	150	64.0	.428	165

Free Agents

PLAYER	POS	AGE	STATUS	2017 TEAM	NEW TEAM	YRS	RK	DOLLARS
Eric Hosmer	1B	28	Signed	Royals	Padres	8	7	$144,000,000
Yu Darvish	SP	31	Signed	Dodgers	Cubs	6	1	$126,000,000
J.D. Martinez	RF	30	Signed	Diamondbacks	Red Sox	5	6	$110,000,000
Lorenzo Cain	CF	32	Signed	Royals	Brewers	5	2	$80,000,000
Jake Arrieta	SP	32	Signed	Cubs	Phillies	3	4	$75,000,000

5

Assumptions and Choices

▷ Fangraphs for statistics
  ○ Thorough, in-depth and transparent
▷ Year Prior To Signing
▷ Hitters Only
  ○ Pitchers are a different world
▷ 2010-2018 data
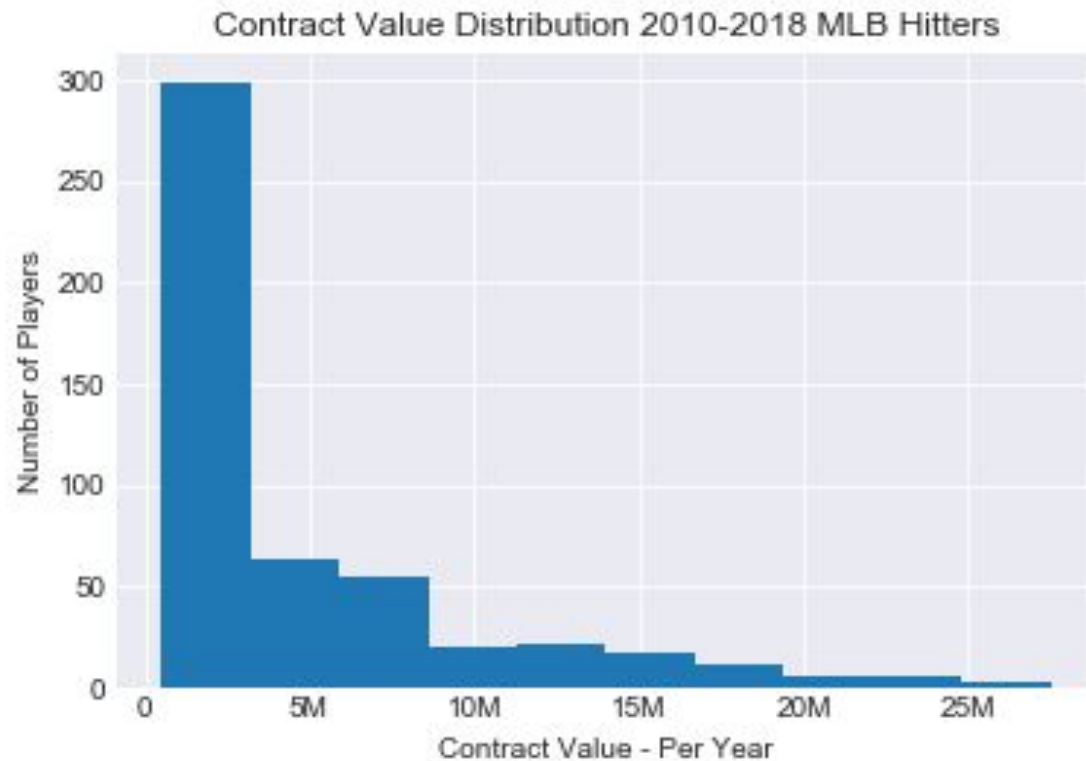 ○ Baseball has changed a lot in the last decade

Assumptions and Choices

▷ Measuring the yearly contract value

20 Mil / 1 Year = 40 Mil / 2 Years

▷ But this value is not normally distributed

  ○ Log transformation

Measurement - Log (Dollars/Year)

Why?

Contract Value Distribution 2010-2018 MLB Hitters
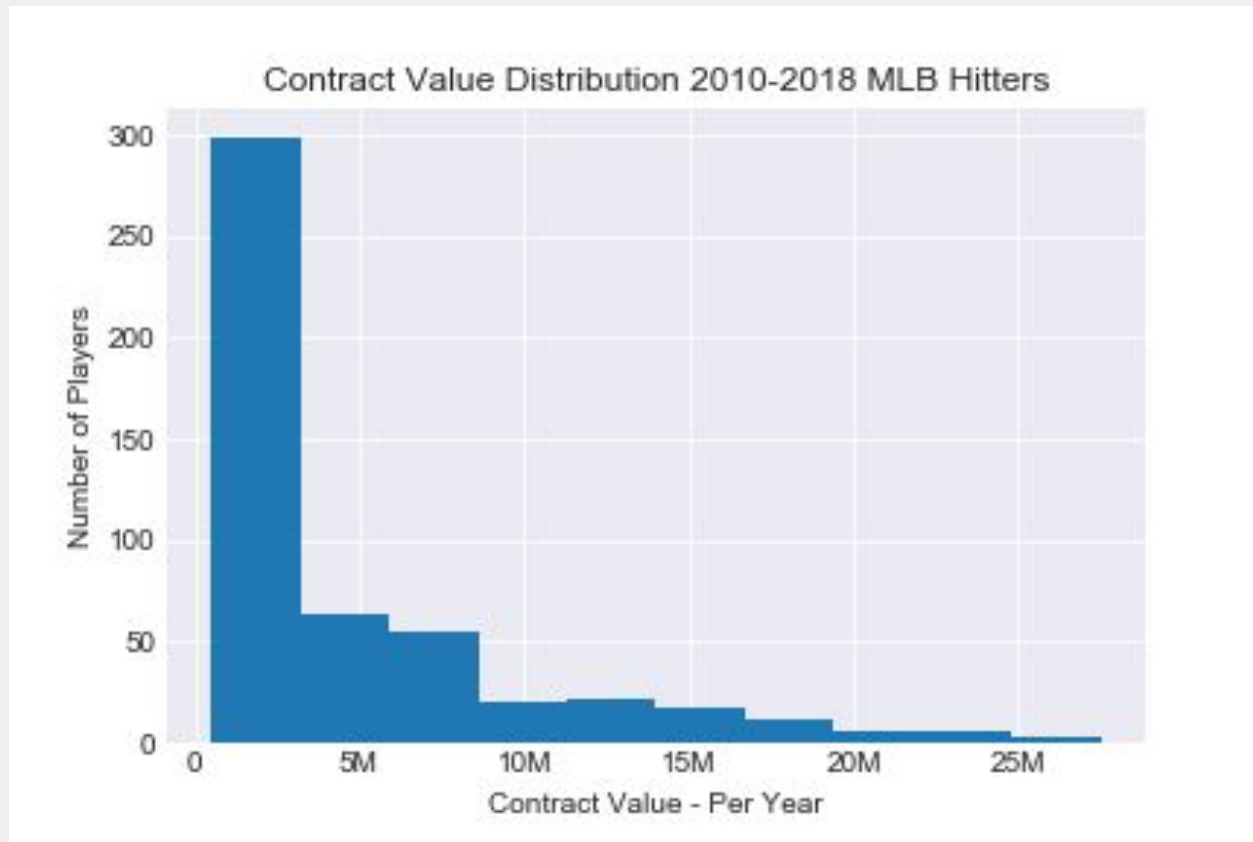
Free Agency

The State of the Market

Free Agency

2017 was a rough year for free agents, especially hitters

Free Agency

Most contracts are for the league minimum or near it

Contract Value Distribution 2010-2018 MLB Hitters

Modeling the Data

Data Pipeline

Web Scraping	Pandas	Linear Regression

- Selenium to automate data gathering across hundreds of pages
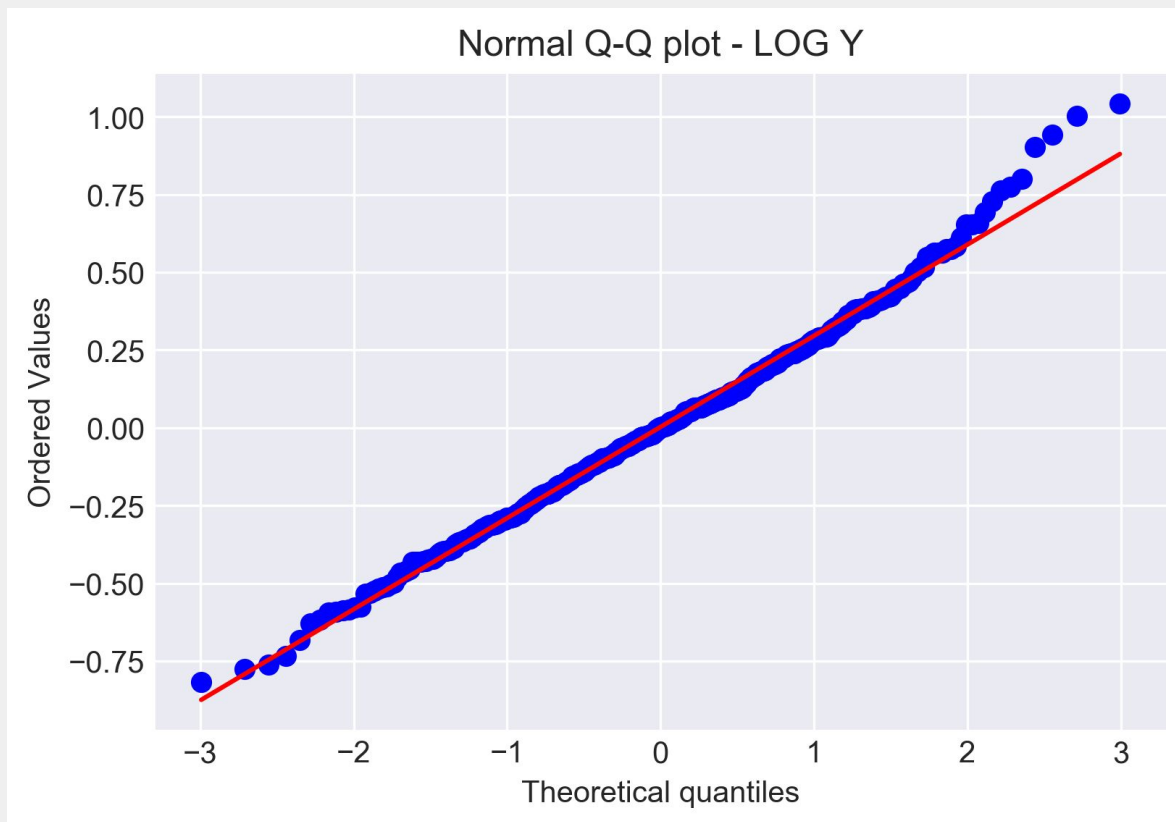- Fangraphs - Hitting Data
- ESPN - Free Agency Data

- Clean Data
- Merge together
- Exploratory Data Analysis

- Data Normalization
- Feature Selection
- Cross-Validation

Measurement - Log (Dollars/Year)

Use Log($/Year)

-Residuals looked much more along the normal expectation

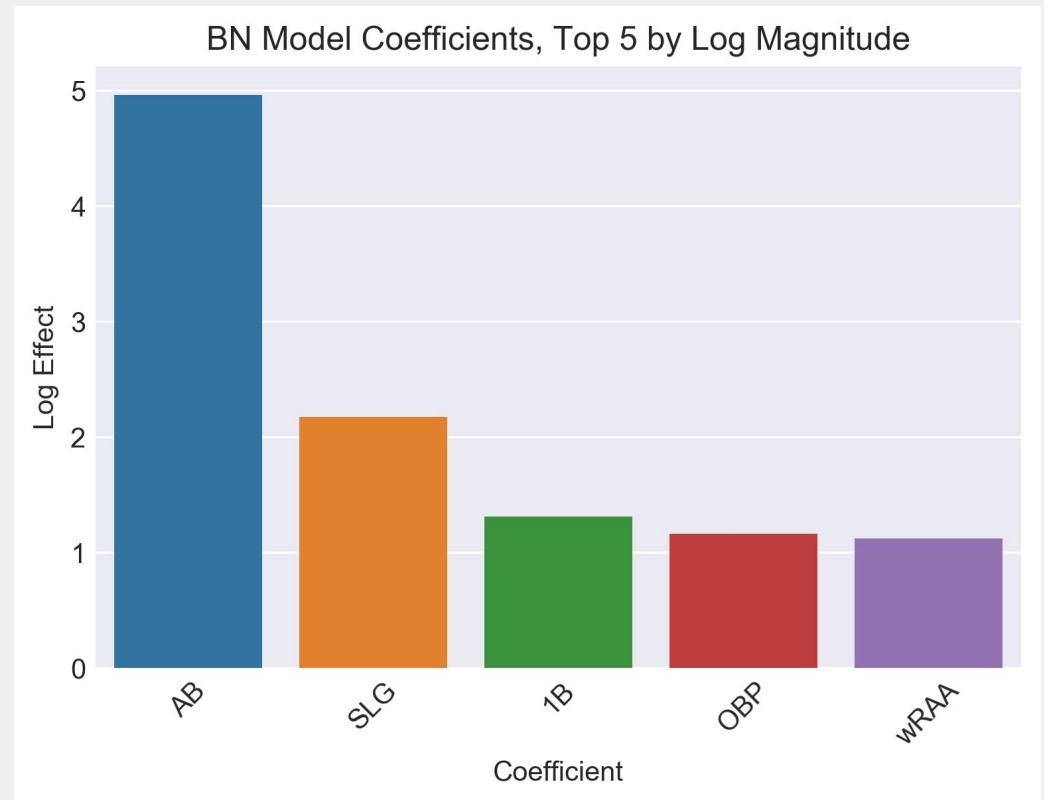Normal Q-Q plot - LOG Y

Final Model

3 Step Pipeline

1. Standard Scaler

2. SelectKBest

   a. 26 Features

3. Linear Regression

Tested Using Cross Validation to find the optimal results on our split data (test set)

Calculated R^2 (on test split) = .6967

BN Model Coefficients, Top 5 by Log Magnitude

So, where does this leave us?

The Client

$15,189,140

Using Conforto's 2017 statistics, our model projects that he <u>would</u> receive just over 15 million dollars on the open market

The Prediction

Reflecting some of our model's limitations

▷ Forecasting for future superstars
▷ Team Preferences/Needs
▷ Bidding Wars

The "human side" of free agency

Blackmon chooses a sure thing over free agency

Brandon Phillips thanks Cincinnati fans for 'unconditional love'

Show me the money!

Thanks!

Any questions?

Brian Newborn

-April 2018

bri.newborn@gmail.com