

Trust the Process: Predicting Philly's NBA Season



David Cortes

TABLE OF CONTENTS

1. Background
2. Problem Statement
3. Data
4. EDA
5. Modeling
6. Predictions
7. Conclusion
8. Next Steps



Background: Sixers in August

The Philadelphia 76ers weren't seen as a contender after the 2019-2020 NBA season.

- Sixth in the Eastern Conference
- First round exit in the playoffs
- Head coach Brett Brown was fired



Background: Sixers Now

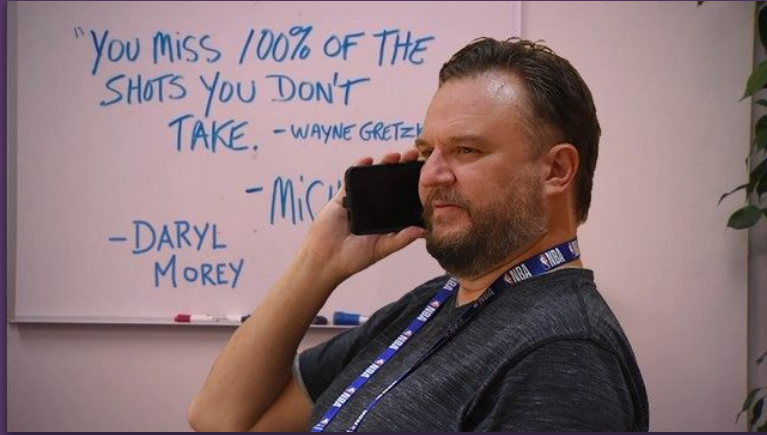
The Sixers look like a different team

- First in the Eastern Conference
- Fourth in the league
- Joel Embiid averages 27.7 points

What changed?



Background: New Front Office



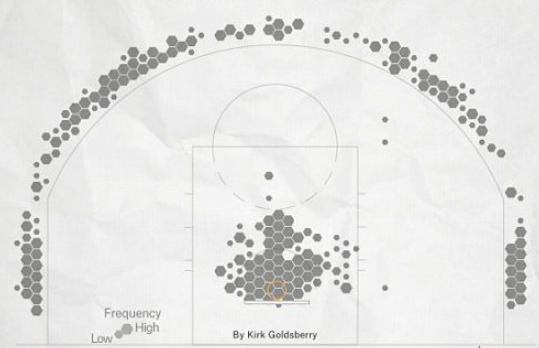
Soon after the end of the 2019-2020 season, 76ers GM Elton Brand hired Daryl Morey to be the new president of basketball operations

- Morey was the Houston Rockets GM for 13 years
- His analytics driven approach changed the way basketball is played in the NBA
- The most efficient spots to take a shot from are just beyond the three point line and right under the basket
- Morey's play style ("Moreyball") takes advantage of these spots
- Rockets broke the single-season record for most three-pointers made during the 2017-2018 season

Overall NBA Points Per Shot
2014-15 Season



HOUSTON ROCKETS
COMMON SHOT LOCATIONS



GRANTLAND





Problem Statement

This capstone aims to build a regression model with 2017-2018 Houston Rockets data in order to predict how many three-point field goals the Philadelphia 76ers make during the 2020-2021 NBA season.

Data

The training and testing data were scraped from Basketball-Reference.com

- Web scraping is the process of collecting data from websites
- Training data: game logs for each player on the 2017-2018 Rockets roster
- Testing data: game logs for each player on the 2020-2021 Sixers roster

The screenshot displays the Basketball-Reference.com website. The top navigation bar includes links for Sports Reference, Baseball, Football (college), Basketball (college), Hockey, F1, Blog, Stathead, and Widgets. A search bar is located on the right. Below the navigation bar, the main content area is divided into several sections:

- Every NBA Player:** A grid of player headshots.
- View any Active Player:** A dropdown menu to select a team, followed by a dropdown to select a player, and a "Go!" button.
- Select a Hall of Famer:** A dropdown menu to select a player.
- Recent Debut:** A list of recent debuts including Anton Harrison (MIN), Naji Marshall (NOP), Ovechia Okongue (ATL), Tyler Bay (DAL), Cassius Winston (WAS), and Jalen Harris (TOR).
- NBA & ABA, WNBA, NBL, G League, and top International players:** A section for international players.
- In Memoriam:** A section for players who have passed away.
- 2020-21 NBA Standings:** A table showing the standings for the 2020-21 NBA season, categorized by East and West.
- Stathead:** A section for advanced statistics and analysis.

The 2020-21 NBA Standings table is as follows:

East	W	L	West	W	L		
IND (1)	58	22	6	LAL (1)	58	24	4
BOS (2)	58	20	6	UTA (2)	58	23	4
MIL (3)	58	20	6	LAC (3)	58	23	5
DEN (4)	58	20	7	ORL (4)	58	20	7
BKN (5)	58	21	8	POR (5)	58	29	7
ATL (6)	58	28	8	MEM (6)	58	27	6
CLE (7)	58	28	9	PHO (7)	58	28	7
ORL (8)	58	28	10	GSW (8)	58	29	8
NOX (9)	58	28	11	SAS (9)	58	29	8
CIN (10)	58	27	10	DAL (10)	58	28	8
CHI (11)	58	27	10	HOU (11)	58	27	9
TOR (12)	58	27	10	OKC (12)	58	27	9
MIA (13)	58	26	10	SAC (13)	58	26	10
DET (14)	58	24	13	NOR (14)	58	25	10
WAS (15)	58	23	10	MIN (15)	58	24	12

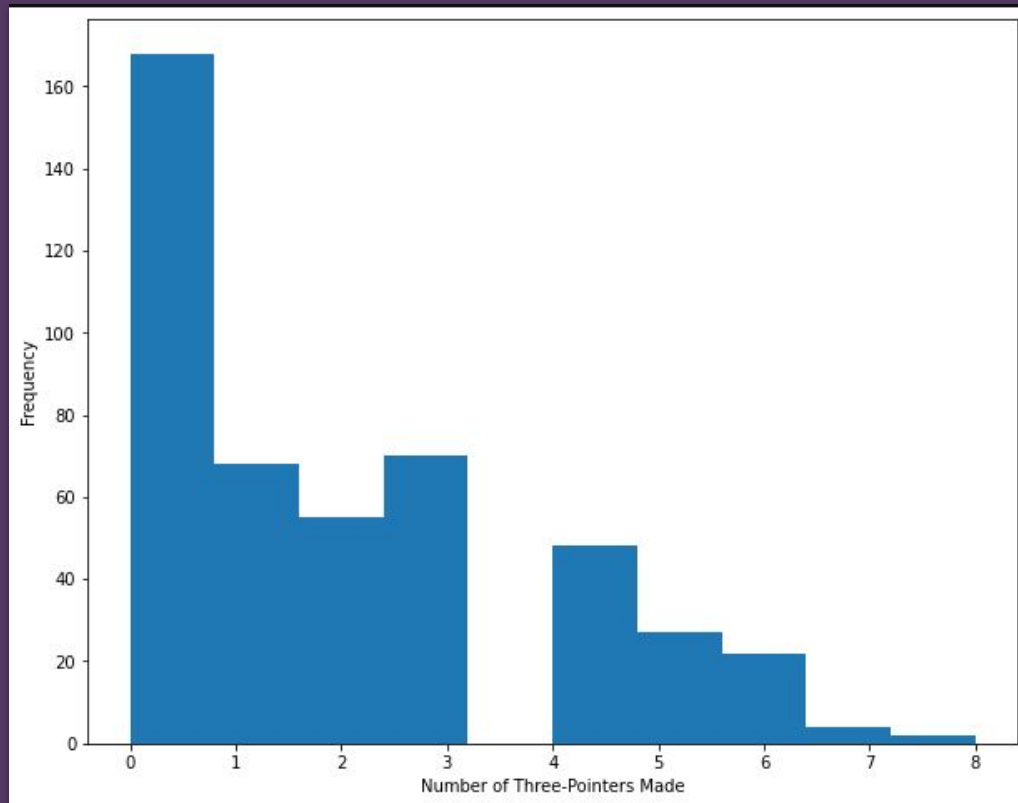
EDA

	HOME/AWAY	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	TRB	AST	PTS	GAME_SCORE	+/-
0	0	5	12	0.417	3	8	0.375	0	0	0.000	8	1	13	11.4	8
1	0	2	7	0.286	2	7	0.286	4	4	1.000	5	0	10	7.4	5
2	1	1	4	0.250	1	4	0.250	0	0	0.000	10	1	3	4.7	19
3	1	2	7	0.286	1	4	0.250	0	0	0.000	6	2	5	3.7	12
4	0	5	11	0.455	4	9	0.444	0	0	0.000	1	0	14	9.3	-1
...
459	1	1	3	0.333	0	0	0.000	0	0	0.000	3	1	2	1.4	13
460	1	4	7	0.571	0	0	0.000	2	2	1.000	7	3	10	15.1	32
461	1	2	5	0.400	0	0	0.000	3	7	0.429	8	0	7	6.5	7
462	1	0	2	0.000	0	0	0.000	2	4	0.500	1	1	2	-1.4	1
463	0	1	4	0.250	0	0	0.000	0	0	0.000	2	0	2	1.2	-1

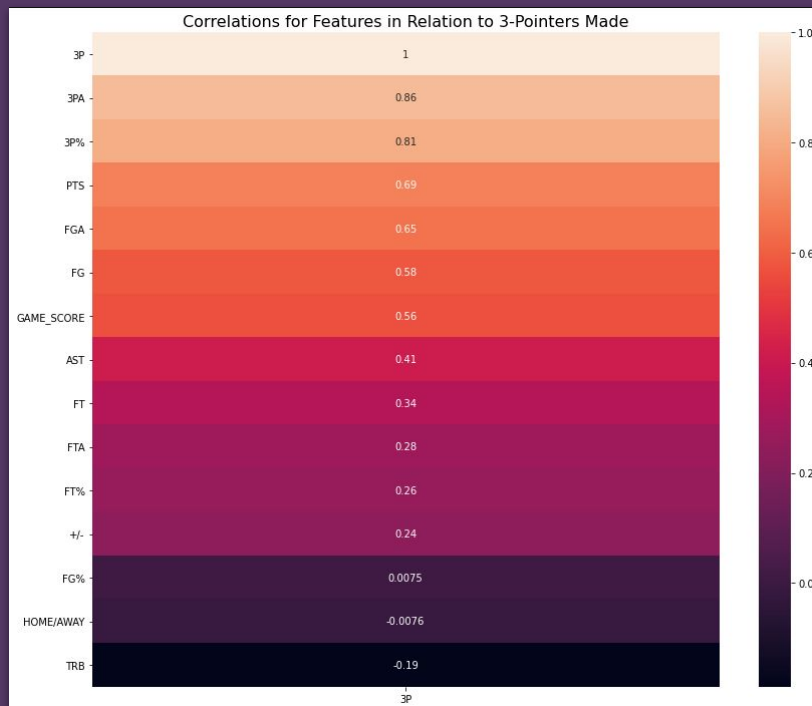
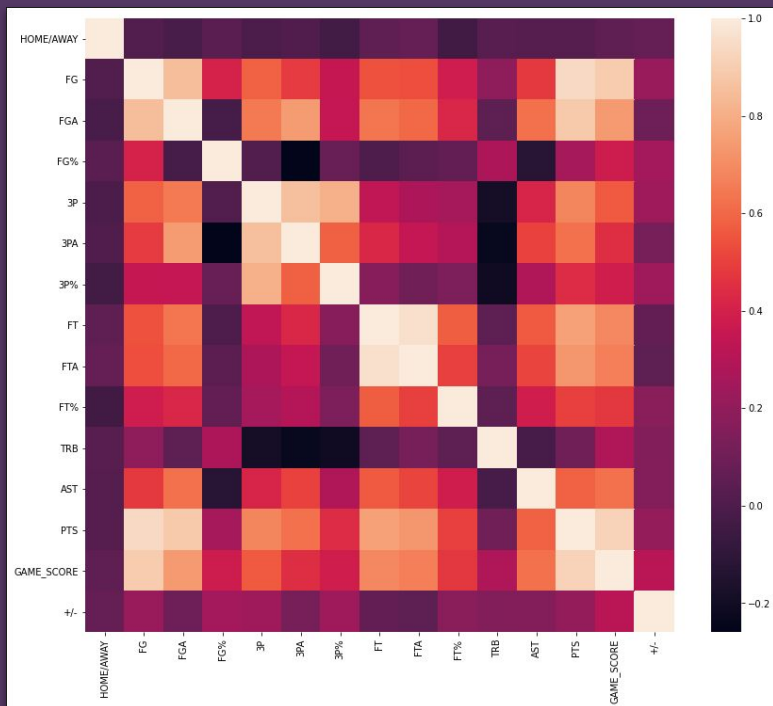
Cleaning and EDA was done on the data once it was gathered

- 'HOME/AWAY' column was binarized
- Null percentage values were converted to zero
- Numerical data was isolated
- Unnecessary columns were dropped

EDA: Frequency of 3-Pointers



EDA: Correlations



Modeling

Linear, Ridge, and Lasso regression models were tested

- PolynomialFeatures was used (degree=2)
- Mean cross-val score:
 - Lasso: 0.79696671
 - Ridge: 0.798621953
 - Linear: 0.8



Lasso Regression

This model returned the best accuracy

Training Score

0.999241585099331

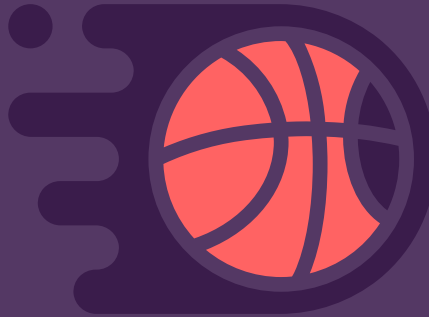
Testing Score

0.997952083858883

Predictions

Three-pointers for the Sixers were predicted using Lasso

- Sum of predicted three-pointers is 164.81
- Since the season just started, there are only 83 rows of data
- Sum of first 83 rows of Rockets 3-pointers is 194



3-Pointers	
0	5.483642
1	0.231284
2	1.355295
3	1.233191
4	0.073508
...	...
78	2.989756
79	0.219648
80	2.823027
81	-0.727994
82	4.084742

Conclusion

- Model turned out very accurate, with a score of over 99%
- Using data from the Sixers' first 18 games, the model was able to predict about 165 three-pointers
- Hopeful that this data and model can be useful for purposes like training, fantasy basketball, etc.



Next Steps

- Find a way to gather data that shows the number of field goals made from right under the basket.
- Regularly update the testing data as the 2020-2021 NBA season continues.
- Feature engineer more advanced stats (3-point attempt rate, turnover percentage, etc.)
- Increase the size of the training data (combine two or three years of Houston Rockets game logs)
- Explore more models like ElasticNet

Thank you!

Questions?

