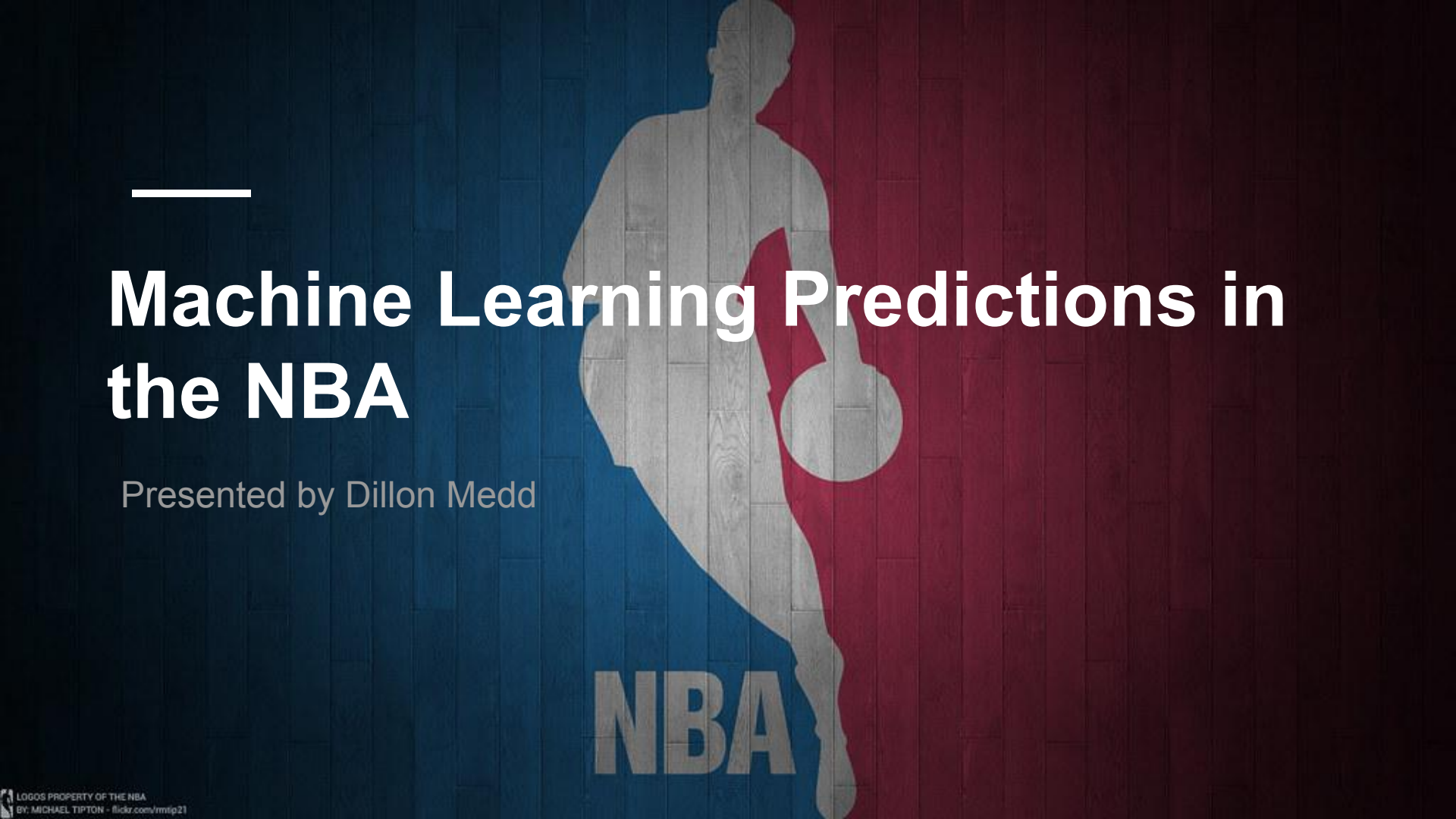


Machine Learning Predictions in the NBA

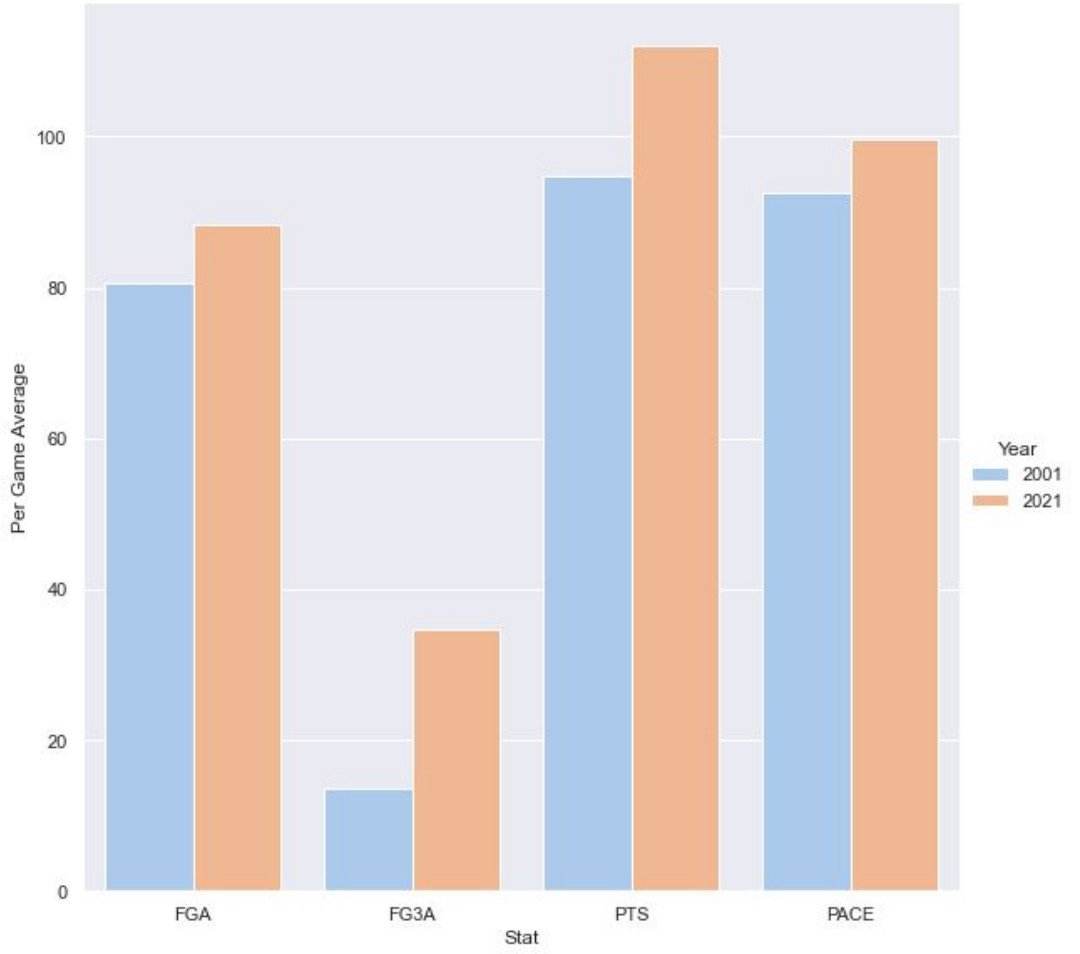
Presented by Dillon Medd

Data and Methods

- Goal: Predict scores of NBA games as accurately as possible
- 4,000 games worth of box score advanced statistics from over over the previous five regular seasons (2017-2021)
  - NBA.com API used to obtain data
- Two separate models were developed
  - Classification model to predict home team result (win/loss)
  - Regression model to predict margin of victory/defeat for home team
- Both models were trained using a combination of lagged team and player statistics

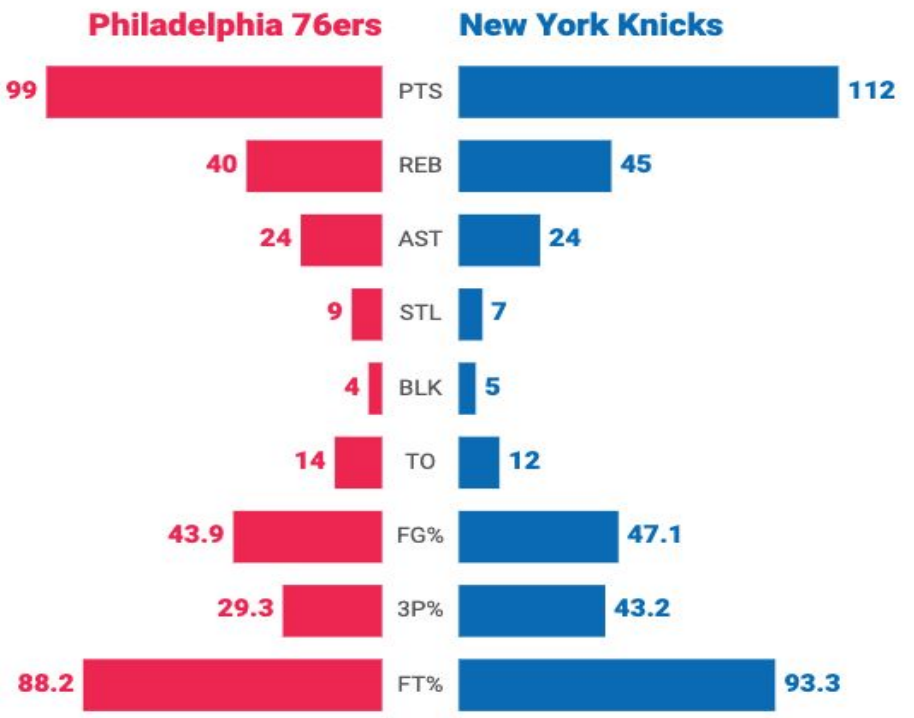
Analytics in the NBA

- “Analytics are part and parcel of virtually everything we do now” — NBA Commissioner Adam Silver
- Advanced statistics led to new strategies and evaluation metrics for teams and players
- Dramatic increase in 3 point attempts, points scored and pace of play in last 20 years



Traditional Box Score

- Traditional Box Scores are easy to follow and understand
- Traditional team and player stats have been tracked since the early days of the NBA
- Simple to calculate
  - Ex. (FG% = Makes / Attempts)



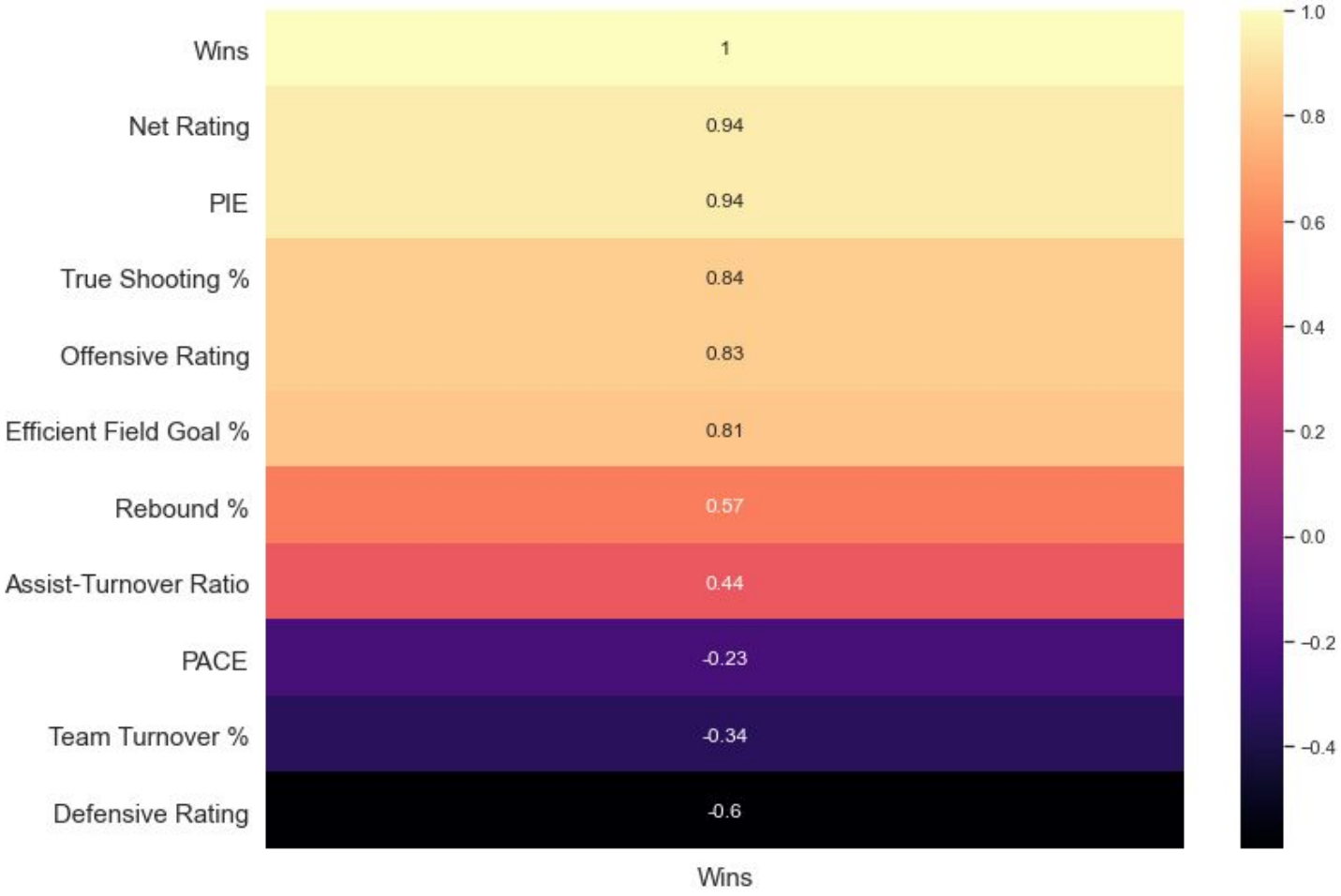
	Basic Box Score Stats																		
Starters	**MP**	**FG**	**FGA**	**FG%**	**3P**	**3PA**	**3P%**	**FT**	**FTA**	**FT%**	**ORB**	**DRB**	**TRB**	**AST**	**STL**	**BLK**	**TOV**	**PF**	**PTS**	**+/-**
Julius Randle	36:13	7	18	.389	2	5	.400	0	0		4	7	11	7	0	0	1	0	16	+8
Evan Fournier	32:12	6	10	.600	4	6	.667	2	2	1.000	1	1	2	4	0	1	0	2	18	+3
RJ Barrett	31:21	4	10	.400	2	4	.500	0	0		0	7	7	1	0	0	4	2	10	-1
Mitchell Robinson	28:59	3	3	1.000	0	0		1	2	.500	3	3	6	1	1	0	1	5	7	0
Kemba Walker	28:10	7	15	.467	5	11	.455	0	0		0	3	3	5	2	1	1	2	19	+3
Team Totals	**240**	**41**	**87**	**.471**	**16**	**37**	**.432**	**14**	**15**	**.933**	**13**	**32**	**45**	**24**	**7**	**5**	**12**	**18**	**112**	

Advanced Box Score

- Advanced Statistics are commonly used by coaches, GMs, writers, etc. to evaluate players and teams in today's NBA
- Provides a better understanding how a player or team played than traditional statistics
- Requires more calculation:
  - Ex. True Shooting Percentage (TS%) = (½) * Points / (FGA + .44 FTA)

	Advanced Box Score Stats															
Starters	**MP**	**TS%**	**eFG%**	**3PAr**	**FTr**	**ORB%**	**DRB%**	**TRB%**	**AST%**	**STL%**	**BLK%**	**TOV%**	**USG%**	**ORtg**	**DRtg**	**BPM**
Julius Randle	36:13	.444	.444	.278	.000	12.9	21.1	17.2	29.2	0.0	0.0	5.3	23.8	120	113	0.4
Evan Fournier	32:12	.827	.800	.600	.200	3.6	3.4	3.5	18.6	0.0	3.6	0.0	15.4	182	118	8.8
RJ Barrett	31:21	.500	.500	.400	.000	0.0	24.4	12.6	4.4	0.0	0.0	28.6	20.3	74	112	-12.8
Mitchell Robinson	28:59	.902	1.000	.000	.667	12.1	11.3	11.7	4.6	1.9	0.0	20.5	7.7	155	113	-3.9
Kemba Walker	28:10	.633	.633	.733	.000	0.0	11.6	6.0	29.3	3.8	4.2	6.3	25.8	136	106	11.2
Team Totals	**240**	**.598**	**.563**	**.425**	**.172**	**31.7**	**72.7**	**52.9**	**58.5**	**7.8**	**12.2**	**11.4**	**100.0**	**125.3**	**110.8**	

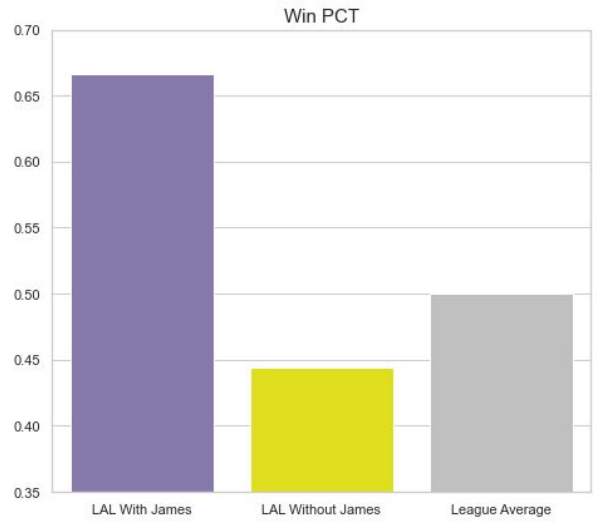
Advanced Statistics Correlation with Win Percentage

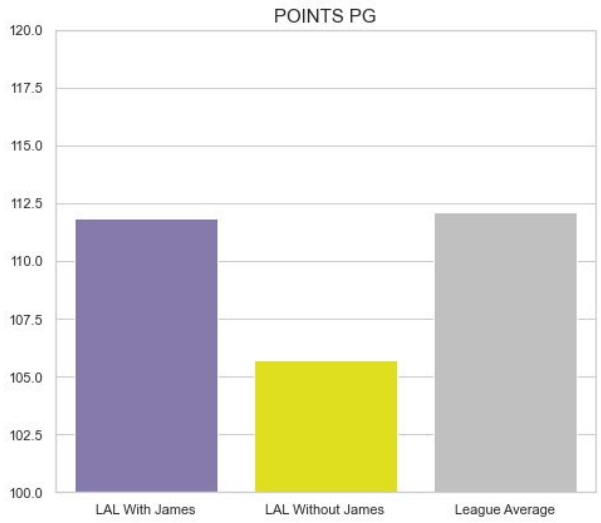


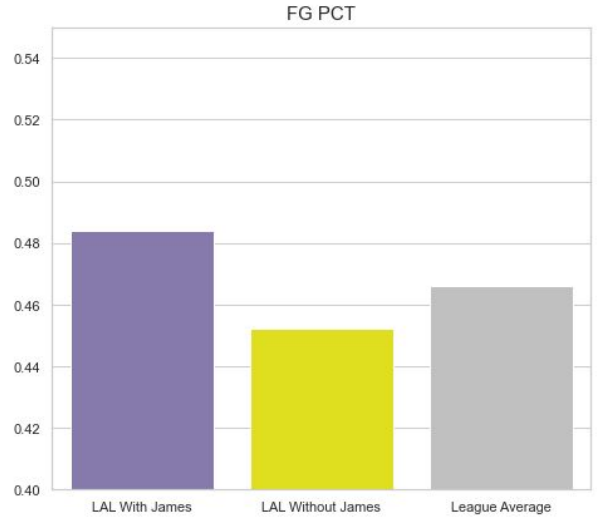
Plotly Graphs

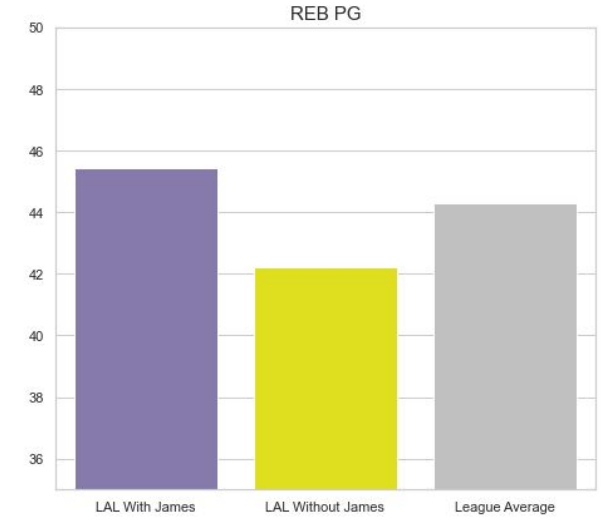
https://chart-studio.plotly.com/~dmedd/7.embed

https://chart-studio.plotly.com/~dmedd/9/#/







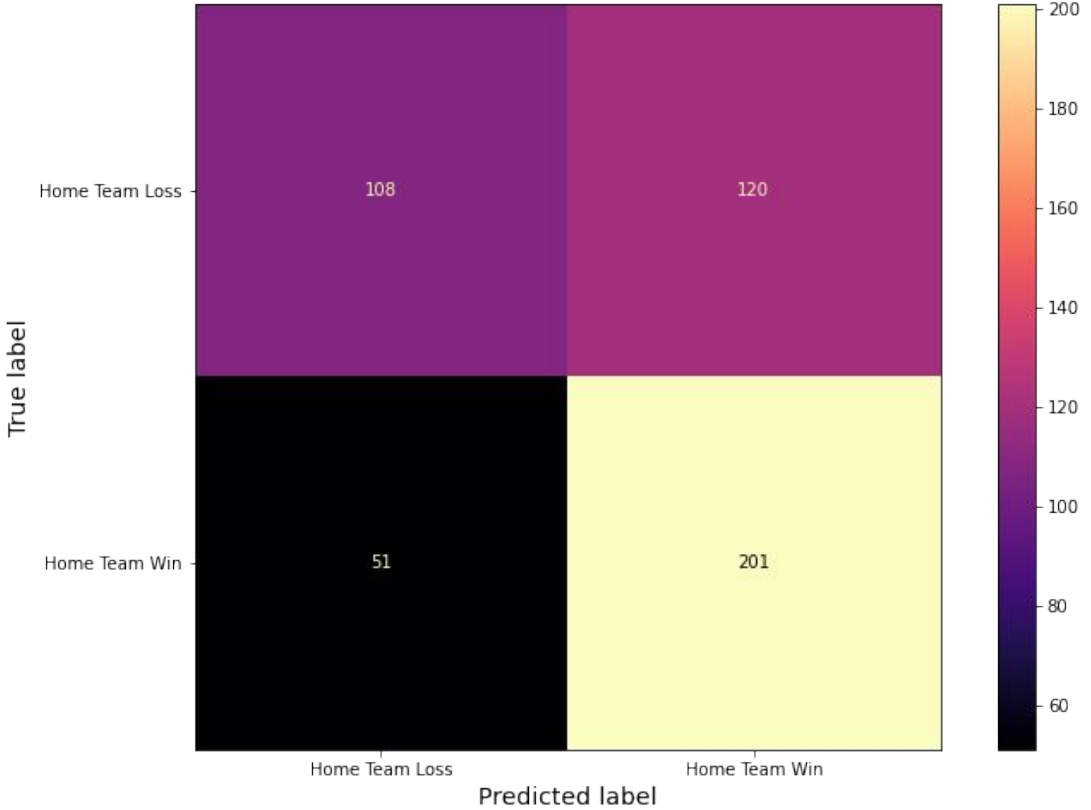


- Today's NBA sees star players rested often and injuries are unavoidable
- Lakers had very different results with and without James in the lineup
- Emphasizes need to include starting lineup information in model


LA FITNESS
LAKERS
23

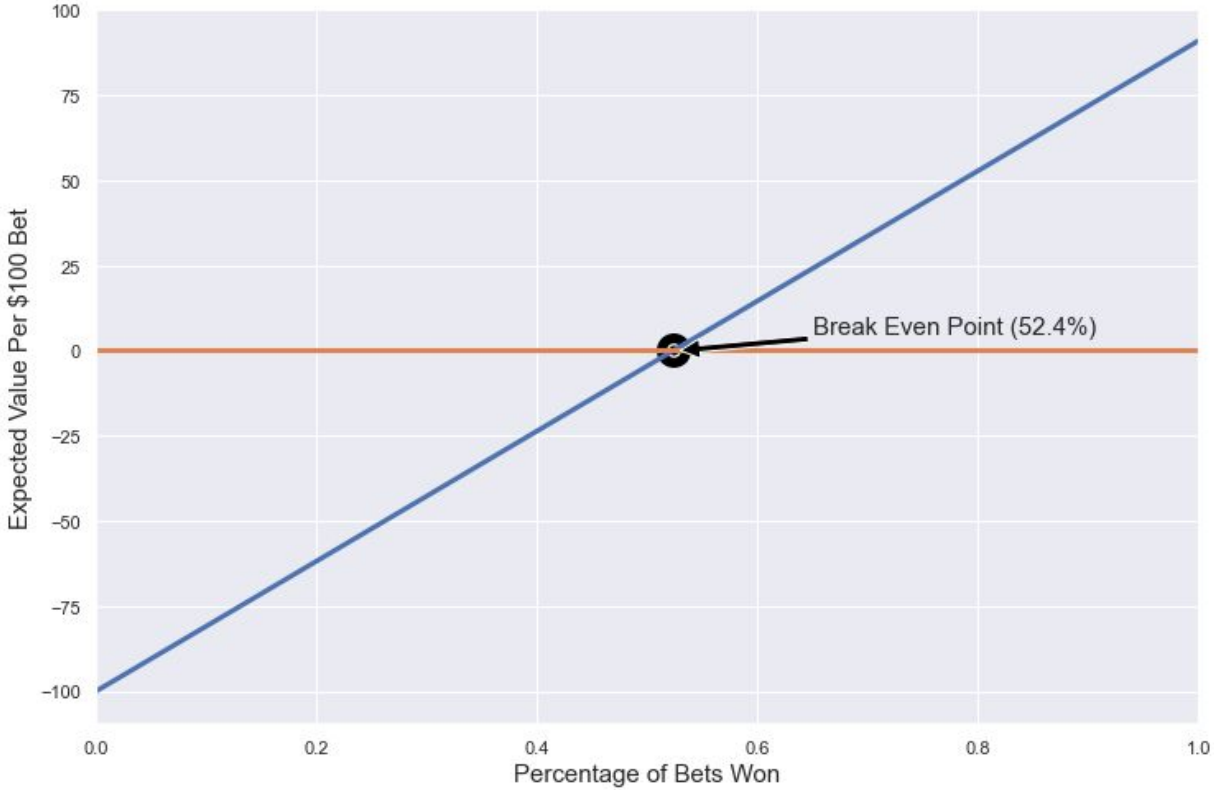
Classification Model Results

- Model predicted winning team correctly for 64.4% of test set
- Logistic Regression Classifier with 22 features
  - 16 Team Statistics
  - 6 Player Statistics
  - Standard Scaler applied



Regression Model Results

- Lasso Regression Model using 28 features returned a RMSE of 12.9
  - 20 Team Statistics
  - 8 Player Statistics
- Best measure of success is to test how it would fare against the ‘spread’
- Due to standard ‘vig’, one must win 52.4% of the time to be profitable
- Model led to a successful bet for 56.1% of test set




Monetary Return of $100 Wagers using Model Predictions
Profit
4000
3000
2000
1000
0
+$3,415
0
100
200
300
400
500
# of Games

Potential Next Steps

- More games, more testing
- Incorporate matchup data into model
- Neural Network
- Test accuracy of classification model predicted probabilities
- Predict exact scores

Thank You!

- dillonmedd1@gmail.com
- GitHub
- LinkedIn