

# Using Statistics to Predict Basketball Games

And Why Dennis Rodman Belongs in  
the Hall of Fame

By Uriel Eckmann



# Project Summary

## Problems

- Can we make inferences about what shapes A game of Basketball?
- Can we predict game winners using previous game Data, and how useful will it be?

## Challenges

- Data Collection



# Workflow

1. Data Collection
  - a. 4 Part WebScrape
2. Data Cleaning and EDA
3. Modeling
  - a. Current Games
  - b. Lagged Games
4. Evaluation
5. Conclusion
6. Recommendations



# Data Collection

## Building A WebScraper of [Basketball-Reference.com](https://www.basketball-reference.com)

Conference Standings * Playoff teams							
Eastern Conference	W	L	W/L%	GB	PS/G	PA/G	SRS
<a href="#">Toronto Raptors*</a>	59	23	.720	—	111.7	103.9	7.29
<a href="#">Boston Celtics*</a>	55	27	.671	4.0	104.0	100.4	3.23
<a href="#">Philadelphia 76ers*</a>	52	30	.634	7.0	109.8	105.3	4.30
<a href="#">Cleveland Cavaliers*</a>	50	32	.610	9.0	110.9	109.9	0.59
<a href="#">Indiana Pacers*</a>	48	34	.585	11.0	105.6	104.2	1.18
<a href="#">Miami Heat*</a>	44	38	.537	15.0	103.4	102.9	0.15
<a href="#">Milwaukee Bucks*</a>	44	38	.537	15.0	106.5	106.8	-0.45
<a href="#">Washington Wizards*</a>	43	39	.524	16.0	106.6	106.0	0.53
<a href="#">Detroit Pistons</a>	39	43	.476	20.0	103.8	103.9	-0.26
<a href="#">Charlotte Hornets</a>	36	46	.439	23.0	108.2	108.0	0.07
<a href="#">New York Knicks</a>	29	53	.354	30.0	104.5	108.0	-3.53
<a href="#">Brooklyn Nets</a>	28	54	.341	31.0	106.6	110.3	-3.67
<a href="#">Chicago Bulls</a>	27	55	.329	32.0	102.9	110.0	-6.84
<a href="#">Orlando Magic</a>	25	57	.305	34.0	103.4	108.2	-4.92
<a href="#">Atlanta Hawks</a>	24	58	.293	35.0	103.4	108.8	-5.30



Regular Season				Share & more ▼	Glossary
G	Date	Start (ET)		Opponent	
1	<a href="#">Thu, Oct 19, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Chicago Bulls</a>	
2	<a href="#">Sat, Oct 21, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Philadelphia 76ers</a>	
3	<a href="#">Mon, Oct 23, 2017</a>	8:30p	<a href="#">Box Score</a>	@	<a href="#">San Antonio Spurs</a>
4	<a href="#">Wed, Oct 25, 2017</a>	10:30p	<a href="#">Box Score</a>	@	<a href="#">Golden State Warriors</a>
5	<a href="#">Fri, Oct 27, 2017</a>	10:30p	<a href="#">Box Score</a>	@	<a href="#">Los Angeles Lakers</a>
6	<a href="#">Mon, Oct 30, 2017</a>	10:00p	<a href="#">Box Score</a>	@	<a href="#">Portland Trail Blazers</a>
7	<a href="#">Wed, Nov 1, 2017</a>	9:00p	<a href="#">Box Score</a>	@	<a href="#">Denver Nuggets</a>
8	<a href="#">Fri, Nov 3, 2017</a>	9:00p	<a href="#">Box Score</a>	@	<a href="#">Utah Jazz</a>
9	<a href="#">Sun, Nov 5, 2017</a>	6:00p	<a href="#">Box Score</a>	<a href="#">Washington Wizards</a>	
10	<a href="#">Tue, Nov 7, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Chicago Bulls</a>	
11	<a href="#">Thu, Nov 9, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">New Orleans Pelicans</a>	
12	<a href="#">Sun, Nov 12, 2017</a>	3:30p	<a href="#">Box Score</a>	@	<a href="#">Boston Celtics</a>

# Data Collection

## Building A WebScraper of [Basketball-Reference.com](https://www.basketball-reference.com)

Regular Season				Share & more ▼	Glossary
G	Date	Start (ET)		Opponent	
1	<a href="#">Thu, Oct 19, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Chicago Bulls</a>	
2	<a href="#">Sat, Oct 21, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Philadelphia 76ers</a>	
3	<a href="#">Mon, Oct 23, 2017</a>	8:30p	<a href="#">Box Score</a>	@ <a href="#">San Antonio Spurs</a>	
4	<a href="#">Wed, Oct 25, 2017</a>	10:30p	<a href="#">Box Score</a>	@ <a href="#">Golden State Warriors</a>	
5	<a href="#">Fri, Oct 27, 2017</a>	10:30p	<a href="#">Box Score</a>	@ <a href="#">Los Angeles Lakers</a>	
6	<a href="#">Mon, Oct 30, 2017</a>	10:00p	<a href="#">Box Score</a>	@ <a href="#">Portland Trail Blazers</a>	
7	<a href="#">Wed, Nov 1, 2017</a>	9:00p	<a href="#">Box Score</a>	@ <a href="#">Denver Nuggets</a>	
8	<a href="#">Fri, Nov 3, 2017</a>	9:00p	<a href="#">Box Score</a>	@ <a href="#">Utah Jazz</a>	
9	<a href="#">Sun, Nov 5, 2017</a>	6:00p	<a href="#">Box Score</a>	<a href="#">Washington Wizards</a>	
10	<a href="#">Tue, Nov 7, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">Chicago Bulls</a>	
11	<a href="#">Thu, Nov 9, 2017</a>	7:30p	<a href="#">Box Score</a>	<a href="#">New Orleans Pelicans</a>	
12	<a href="#">Sun, Nov 12, 2017</a>	3:30p	<a href="#">Box Score</a>	@ <a href="#">Boston Celtics</a>	



Chicago Bulls (0-1)

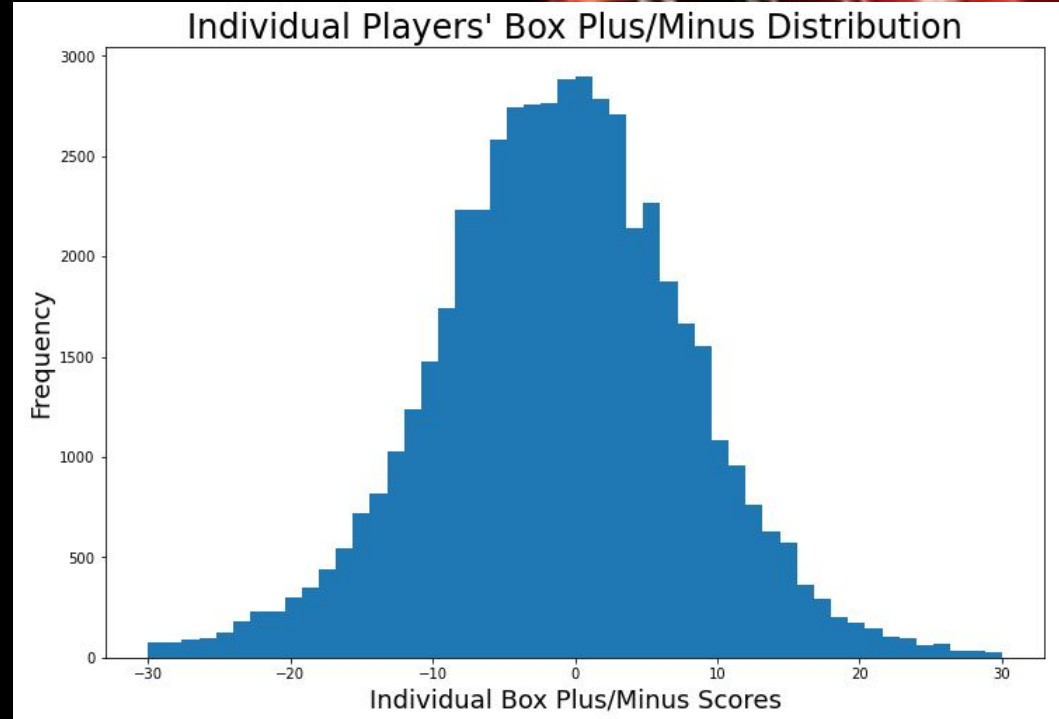
Share & more ▼

Glossary

	Basic Box Score Stats																			
Starters	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	+/-
Justin Holiday	34:53	5	16	.313	4	12	.333	1	1	1.000	0	4	4	2	1	1	1	0	15	+1
Lauri Markkanen	33:27	5	12	.417	2	6	.333	5	6	.833	2	6	8	0	0	1	3	2	17	-2
Robin Lopez	31:44	7	15	.467	0	1	.000	4	4	1.000	4	4	8	2	1	0	1	3	18	+3
Jerian Grant	27:10	3	9	.333	0	1	.000	1	1	1.000	1	1	2	7	2	0	2	1	7	-11
Paul Zipser	24:07	3	6	.500	0	1	.000	1	1	1.000	0	5	5	3	1	0	3	4	7	-15
Reserves	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	+/-
Denzel Valentine	25:31	4	10	.400	4	7	.571	0	0		0	2	2	0	0	0	1	4	12	-19
Cristiano Felício	16:16	2	2	1.000	0	0		0	0		1	1	2	0	0	0	3	4	4	-20
Kay Felder	15:31	2	9	.222	0	1	.000	2	2	1.000	0	2	2	6	0	0	4	1	6	+3
David Nwaba	12:53	1	1	1.000	0	0		2	3	.667	0	4	4	1	0	0	0	0	4	-15
Quincy Pondexter	11:55	4	6	.667	0	1	.000	0	0		2	0	2	0	2	0	1	3	8	-5
Ryan Arcidiacono	5:19	0	2	.000	0	2	.000	0	0		0	0	0	1	0	0	0	1	0	-9
Antonio Blakeney	1:14	1	1	1.000	0	0		0	0		0	0	0	0	0	0	0	0	2	+4
Bobby Portis	Player Suspended																			
Team Totals	240	37	89	.416	10	32	.313	16	18	.889	10	29	39	22	7	2	19	23	100	

# Data Cleaning and EDA

- Redundant Columns
- Dividing by Zero
- Offensive Rating vs. Defensive Rating
- Box Plus/Minus
- Adjusted Box Plus/Minus



# Modeling: Current Games

- Aggregate by Teams
- Reorganize by Teams and Opponents
- Swap Positions
- Logistic Regression
- Decision Tree
- Bagged Decision Tree
- Random Forest Model
- Ada Boost Classifier



# Evaluating: Current Games

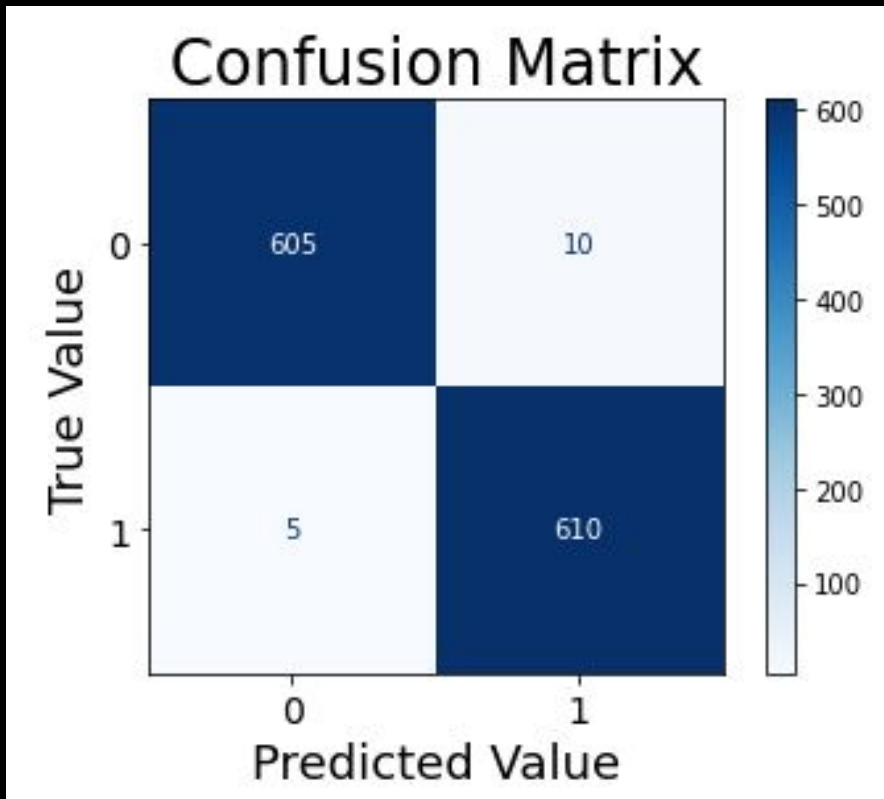
## Logistic Regression

- Train Accuracy: 99.5%
- Test Accuracy: 98.8%
- Cross Validation Mean: 99.1%
- Model is Performing Excellently
- Caveat: This Game is Over

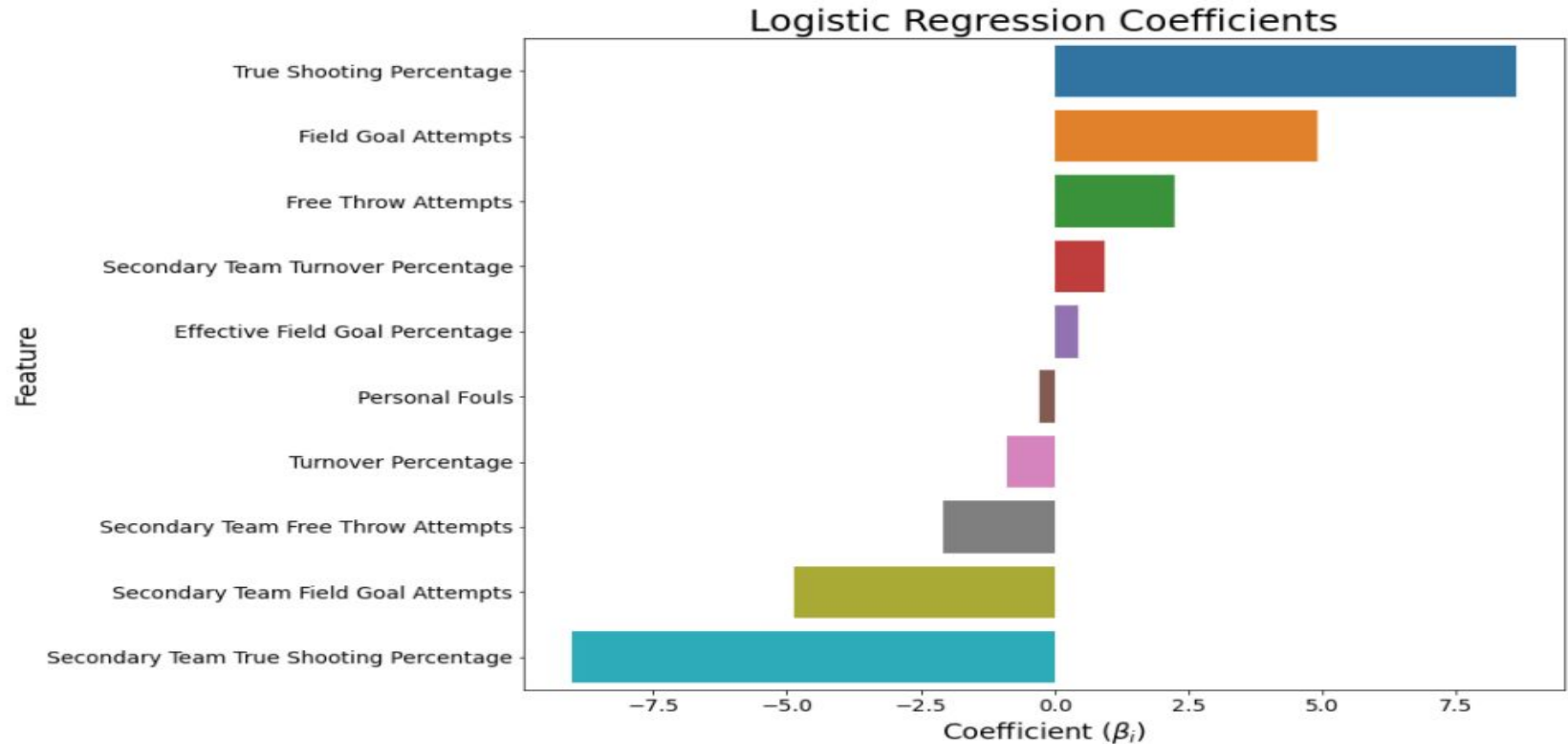




# Evaluating: Logistic Regression



# Evaluating: Logistic Regression



# Modeling: Lagged Games

- Aggregate by Teams
- Lag Each Team
- Reorganize by Teams and Opponents
- Swap Positions
- Logistic Regression
- Decision Tree
- Bagged Decision Tree
- Random Forest Model
- Ada Boost Classifier



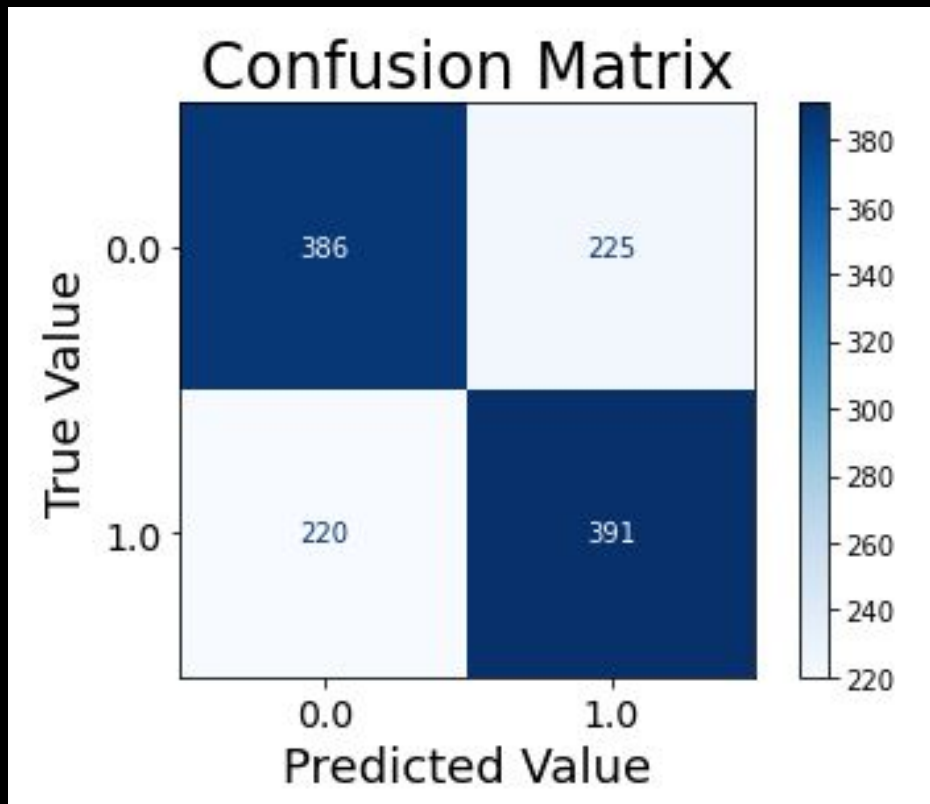
# Evaluating: Lagged Games

## Logistic Regression

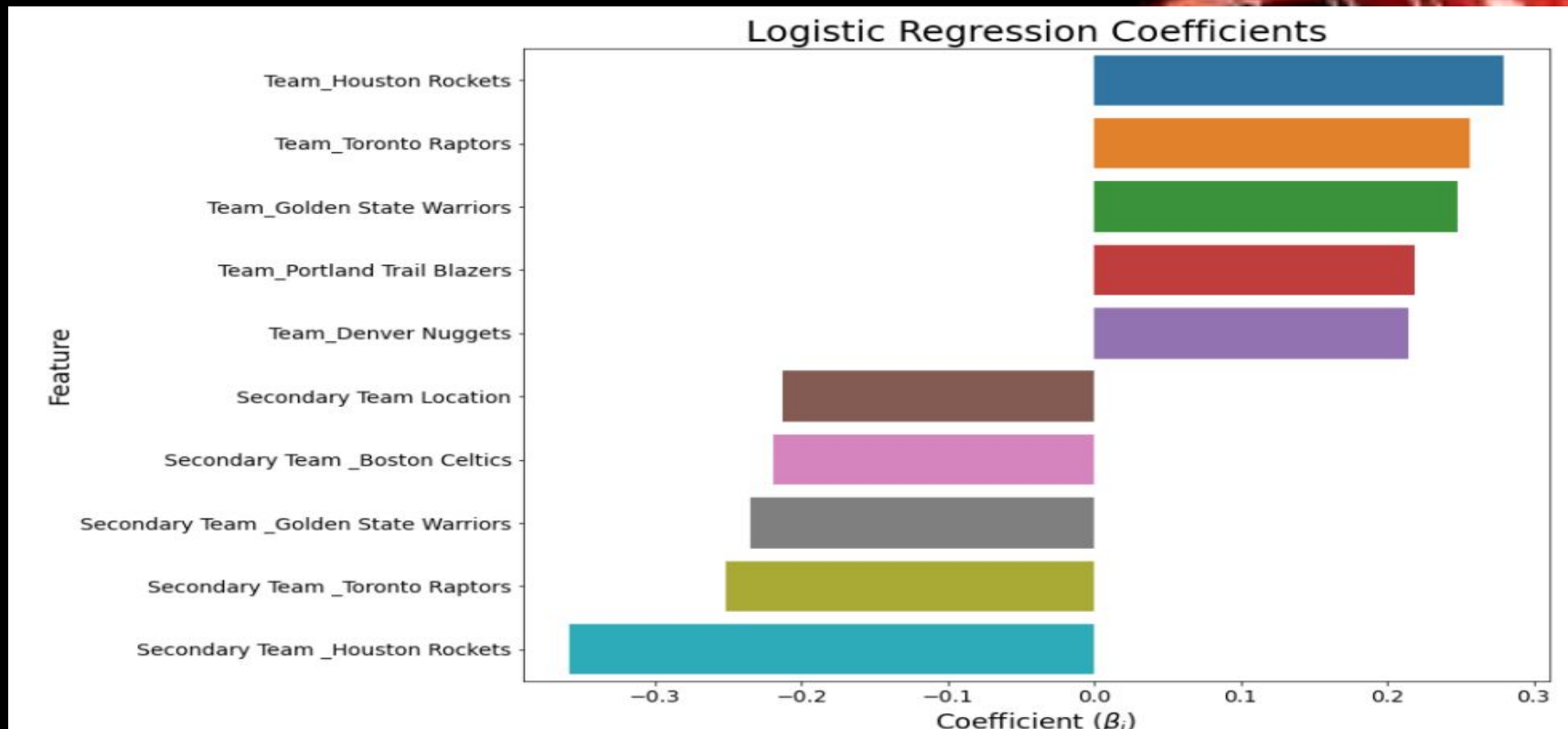
- Train Accuracy: 67.6%
- Test Accuracy: 63.6%
- Cross Validation Mean: 65.8%
- Model is Performing Moderately Well



# Evaluating: Logistic Regression



# Evaluating: Logistic Regression



# Conclusions And Recommendations

- Inferences are Possible
- Predictions are Possible
- Utilize More Data - Lags
- More Models are Needed
- Slim Down Certain Features
- Build Regression Models

