Daryl Morey Group

Andrew Cannon, Bradley Brown, Chris Harden, Felicia Seng,
Isaac Corcoran, Jordan Khamvongsouk

Agenda

Project Overview
Overview of Project and data sources

Modeling and Variable Selection
Models Used to Predict NBA Team Success

Conclusion
Findings From Analysis and Model

Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation

Agenda

Project Overview
Overview of Project and data sources

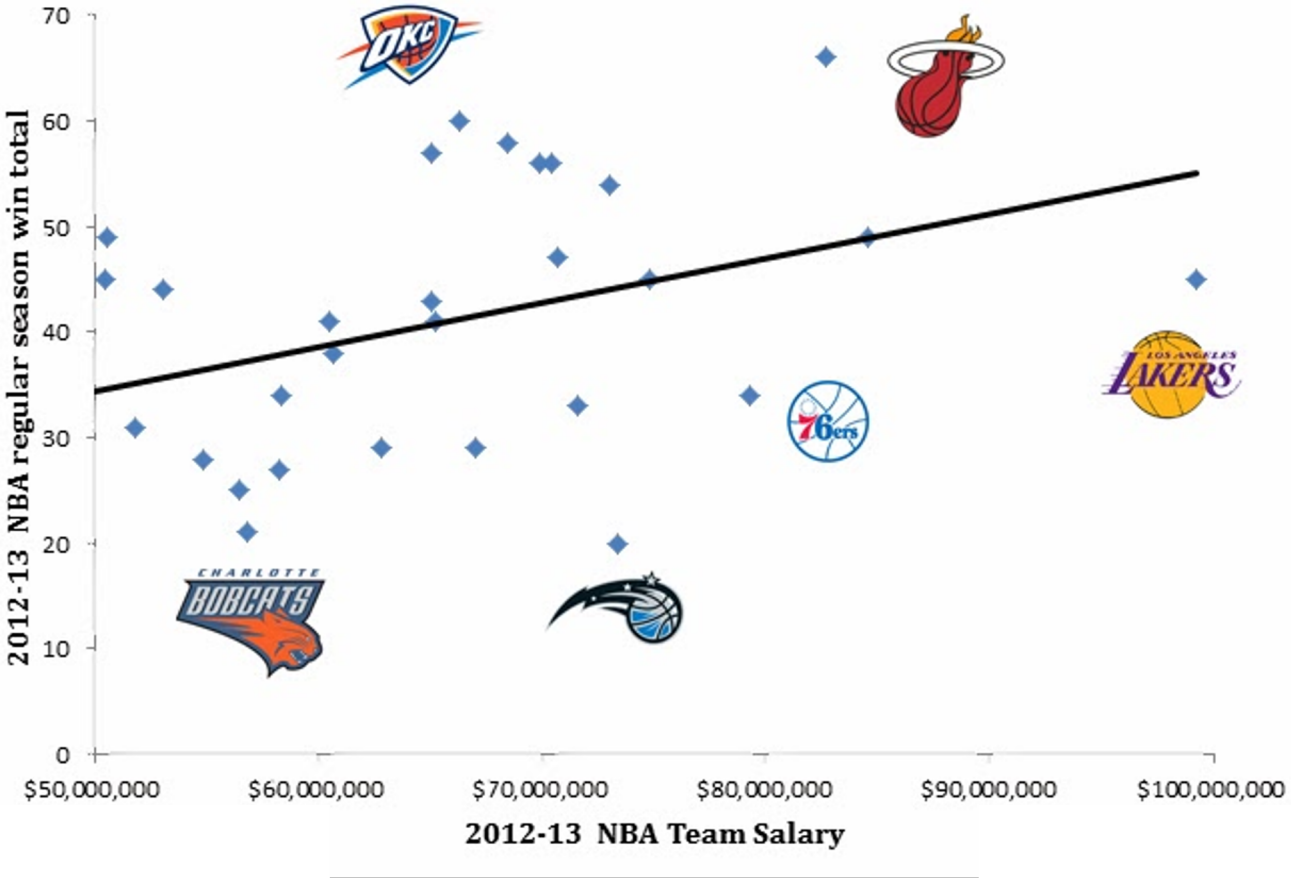
Modeling and Variable Selection
Models Used to Predict NBA Team Success

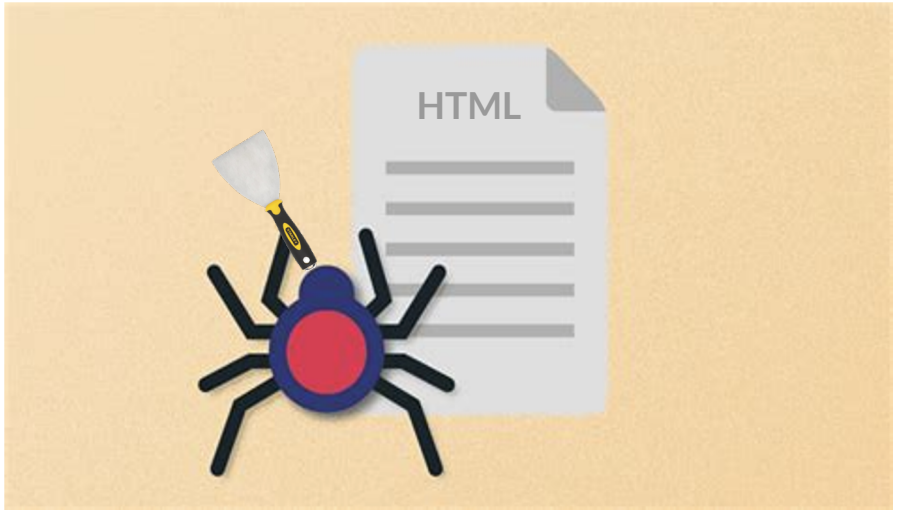
Conclusion
Findings From Analysis and Model

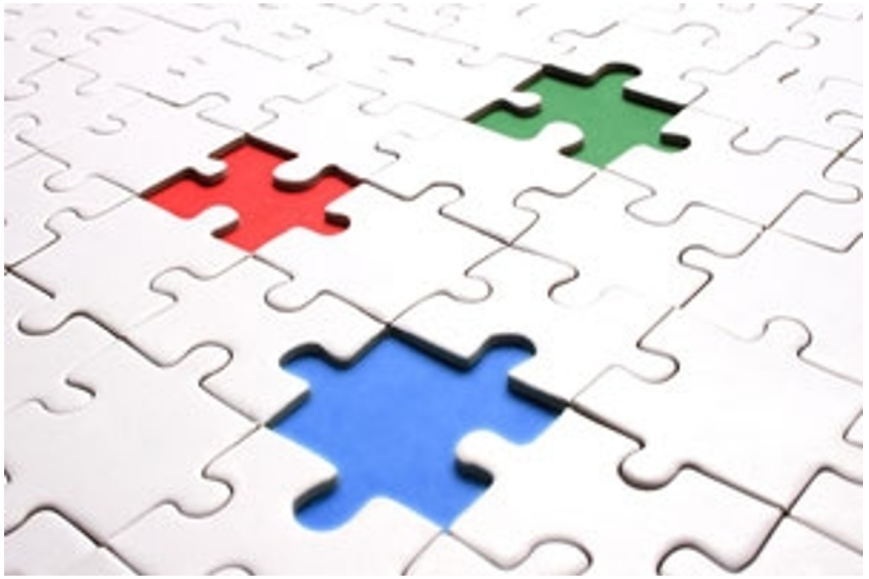
Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation


INJURY
RESULTS
TRAINING
XBOX
70%
85%
60%
90%


2012-13 NBA regular season win total
70
60
50
40
30
20
10
0
$50,000,000
$60,000,000
$70,000,000
$80,000,000
$90,000,000
$100,000,000
2012-13 NBA Team Salary
OKC
CHARLOTTE BOBCATS
76ers
LOS ANGELES LAKERS


HTML



ELO	EQUIVALENT RECORD	TEAM DESCRIPTION
1800	67-15	All-time great
1700	60-22	Title contender
1600	51-31	Playoff bound
1500	**41-41**	**Average**
1400	31-51	In the lottery
1300	22-60	LOL
1200	15-67	Historically awful

Agenda

Project Overview
Overview of Project and data sources

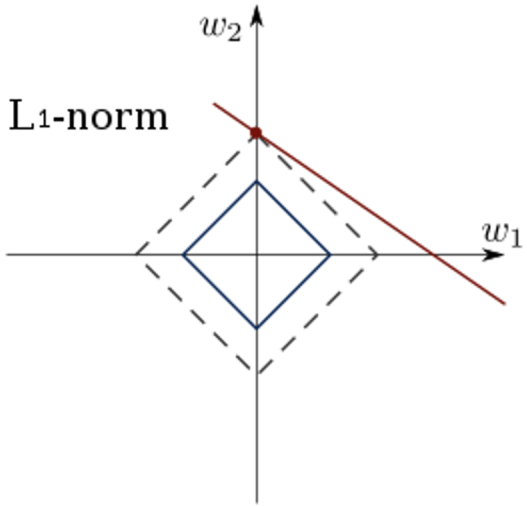
Modeling and Variable Selection
Models Used to Predict NBA Team Success

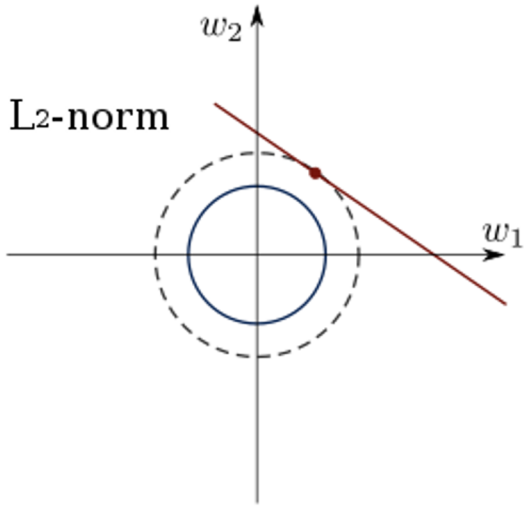
Conclusion
Findings From Analysis and Model

Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation

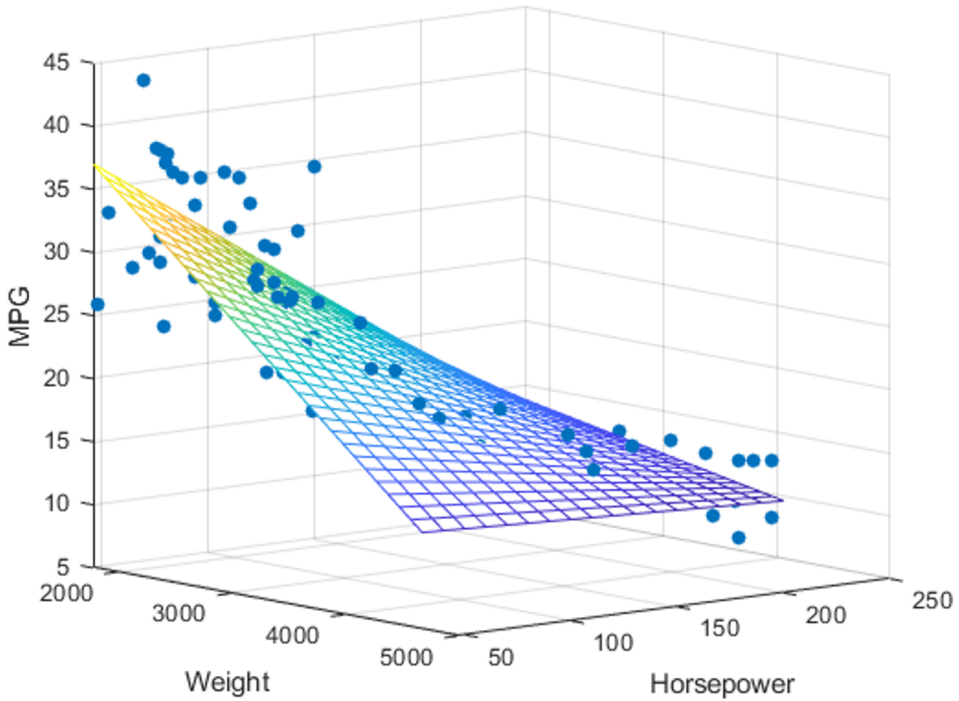
Variable Selection





1. Lasso
2. Ridge Regression
3. Elastic NET

Linear Model



Benefits:

- Works with almost any kind of data
- Very informative and interpretable models

Downsides:

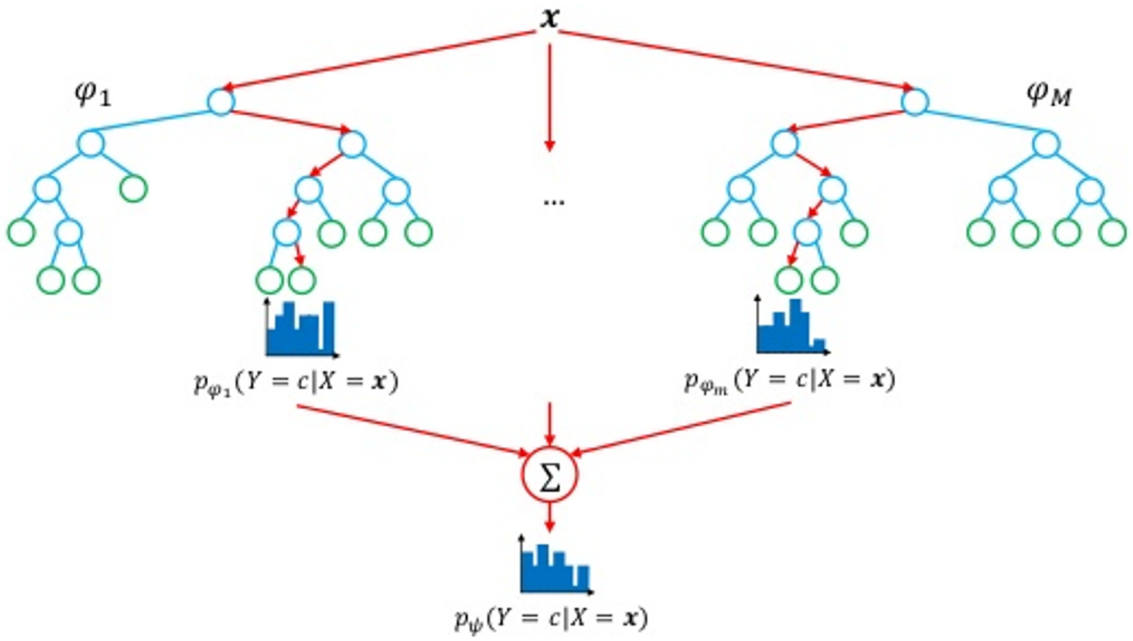
- Many assumptions have to met

Results

	id	predicted	actual	diff
329	POR-2003	1712.302	1631.086	-81.2161
581	LAL-2013	1686.209	1569.041	-117.1681
500	LAL-2010	1662.924	1724.231	61.3072
490	BOS-2010	1656.463	1696.829	40.3656
250	SAS-2000	1643.973	1723.002	79.0293
488	ATL-2010	1641.171	1635.187	-5.9842
120	OKC-1995	1639.103	1717.171	78.0680
330	SAC-2003	1635.517	1722.131	86.6139
522	DAL-2011	1628.768	1735.933	107.1649
182	WAS-1997	1628.684	1567.351	-61.3335
32	CHI-1992	1626.167	1782.122	155.9554
309	BKN-2003	1625.398	1664.327	38.9287
273	PHO-2001	1623.179	1630.131	6.9517
466	DAL-2009	1617.980	1635.159	17.1789
517	ATL-2011	1616.502	1608.431	-8.0710

- Predictions for max Elo were within a reasonable range (+/- 75) of the actual max Elo for a team 73.44% of the time
- The most egregious prediction was the 2010 New York Knicks, when the model was off by 259 points. Predicted: 1215 (Historically Awful), Actual: 1474 (Average Team)
- The most accurate prediction was the 1995 Denver Nuggets, where the model predicted 1596.4 and the actual was 1597.4.
- The model considered the following variables to be significant, listed by order of significance:
  - tot_spVORP
  - tot_spWS
  - min_spPTS
  - std_sal
  - tot_sal
  - min_spPER

Random forests



Randomization

- Bootstrap samples
- Random selection of $K \leqslant p$ split variables
- Random selection of the threshold

} Random Forests

} Extra-Trees

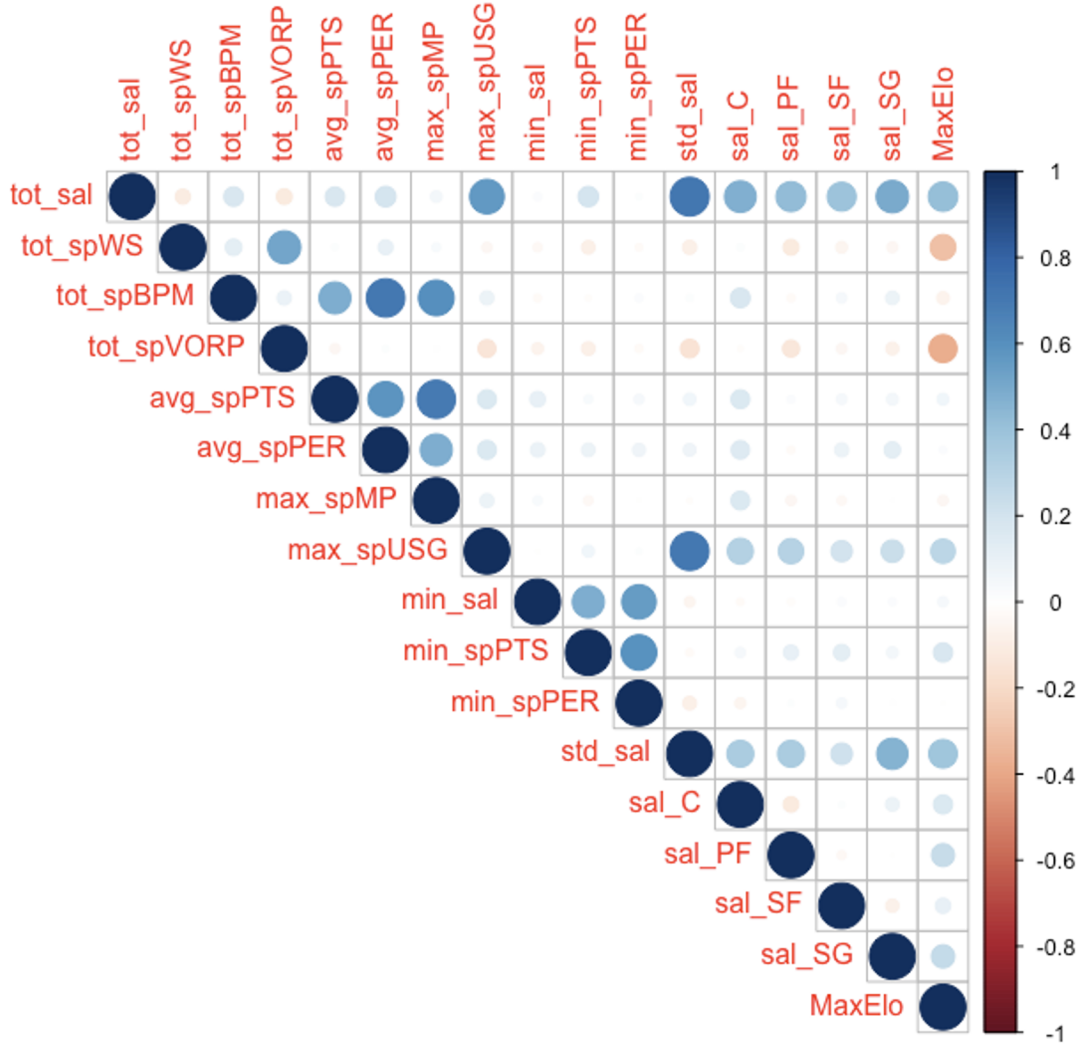
Benefits:

- Reduces overfitting
- Improves accuracy
- Normalized data is not necessary

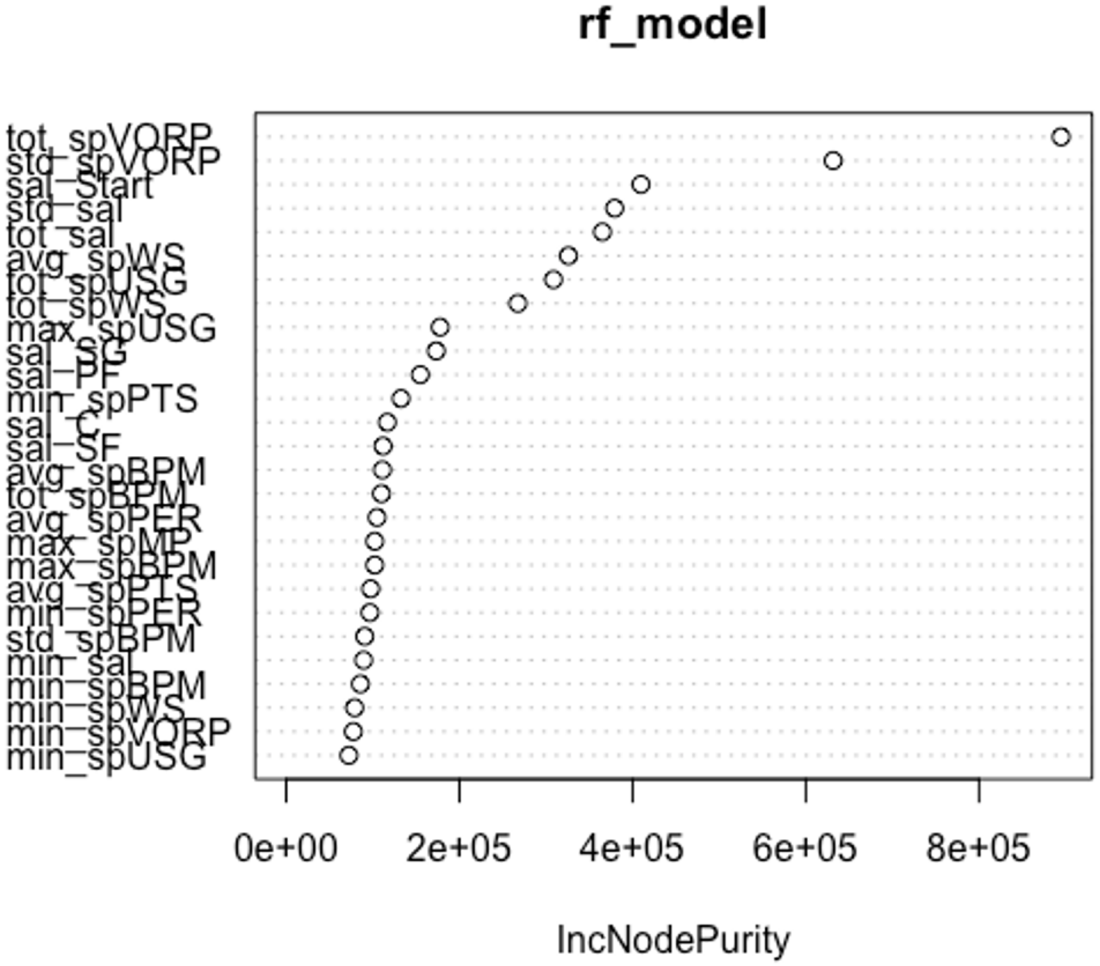
Downsides:

- Less interpretability
- Harder to establish variable importance

Random Forest Model Variables



Random Forest Model Variables



Results

	id	predicted	actual	diff
484	SAS–2009	1689.705	1666.912	–22.7933
577	DEN–2013	1670.144	1697.675	27.5308
515	UTA–2010	1668.926	1666.756	–2.1704
243	LAL–2000	1668.501	1779.287	110.7859
522	DAL–2011	1665.425	1735.933	70.5079
126	UTA–1995	1660.554	1710.618	50.0640
490	BOS–2010	1658.453	1696.829	38.3756
250	SAS–2000	1654.549	1723.002	68.4533
395	GSW–2006	1653.356	1535.942	–117.4140
93	MIA–1994	1652.412	1617.463	–34.9492
581	LAL–2013	1651.950	1569.041	–82.9091
52	UTA–1992	1651.760	1659.517	7.7572
330	SAC–2003	1650.102	1722.131	72.0289
611	MEM–2014	1649.783	1632.341	–17.4418
500	LAL–2010	1647.449	1724.231	76.7822

- Predictions for max Elo were within a reasonable range (+/- 75) of the actual max Elo for a team 85.94% of the time
- The most egregious prediction was the 2004 San Antonio Spurs, when the model was off by 213 points. Predicted: 1551 (Playoff Bound), Actual: 1764 (Title Contender)
- The most accurate prediction was the 2008 Golden State Warriors, where the model predicted 1623.14 and the actual was 1623.10.
- The model considered the following variables to be most important, listed in order:
  - tot_spVORP
  - std_spVORP
  - sal_Start
  - std_sal
  - tot_sal
  - avg_spWS

Agenda

Project Overview
Overview of Project and data sources

Modeling and Variable Selection
Models Used to Predict NBA Team Success

Conclusion
Findings From Analysis and Model

Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation

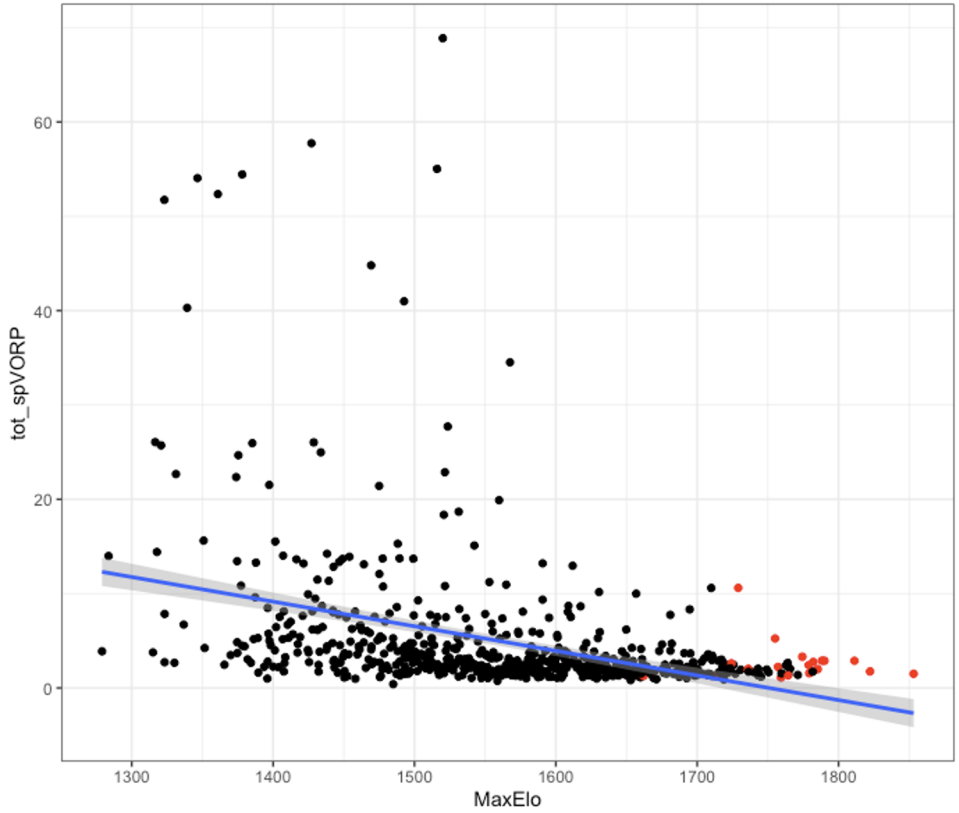
Conclusions

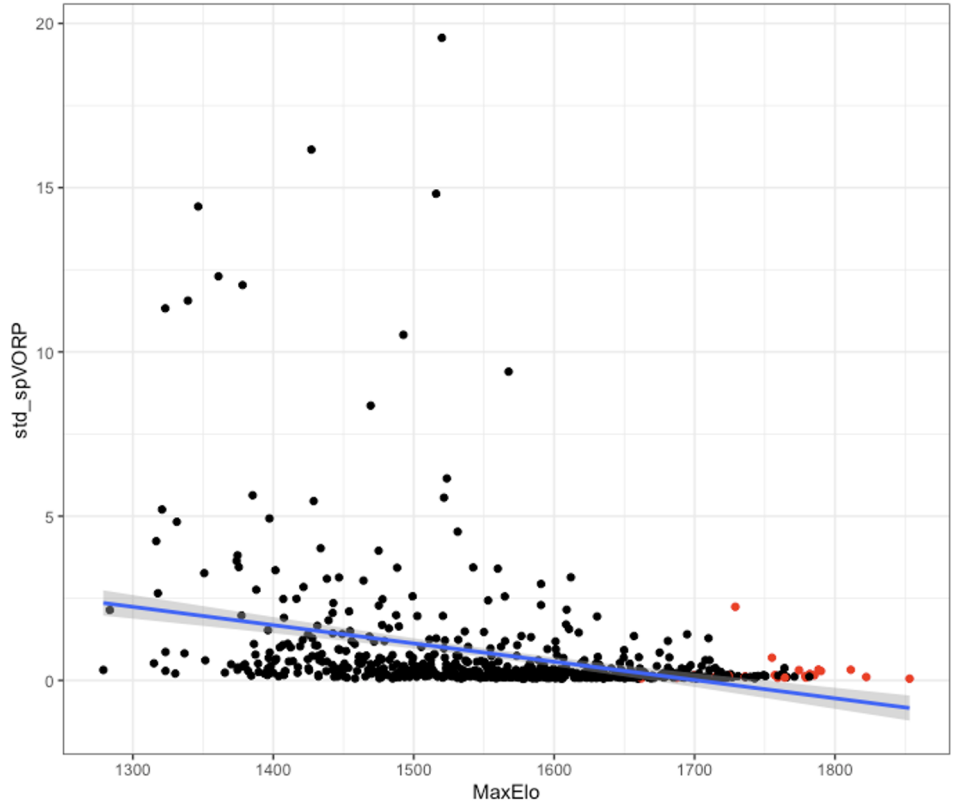
- Our model is quite good at accurately predicting the category a team will fall into (Lottery Bound, Average Team, Playoff Bound, Title Contender, etc.)

- As a team spends less of its salary cap for player production, especially measured by VORP, they tend to perform better during the season, specifically having a higher maximum ELO rating.

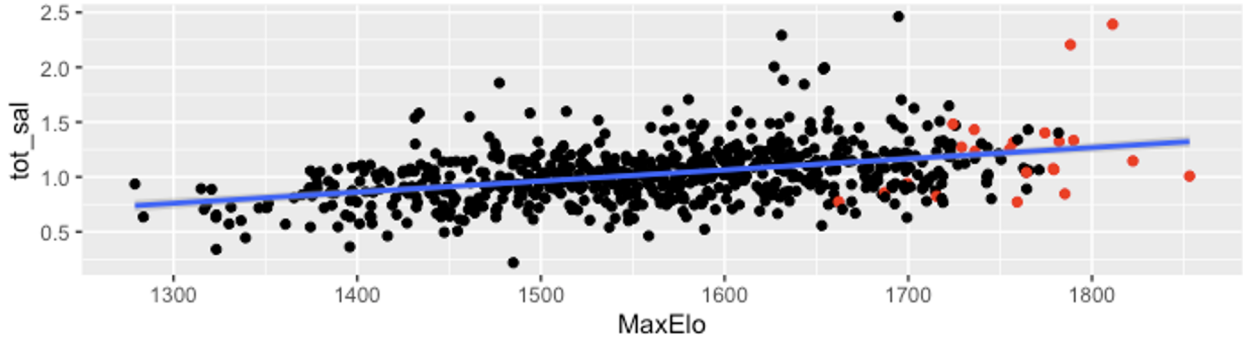
- Paying for superstars (having large contracts) correlates with having a higher max ELO, especially if you're able to pay less for other players too, creating a higher standard deviation. Of course, just paying a lot for a player doesn't guarantee success.

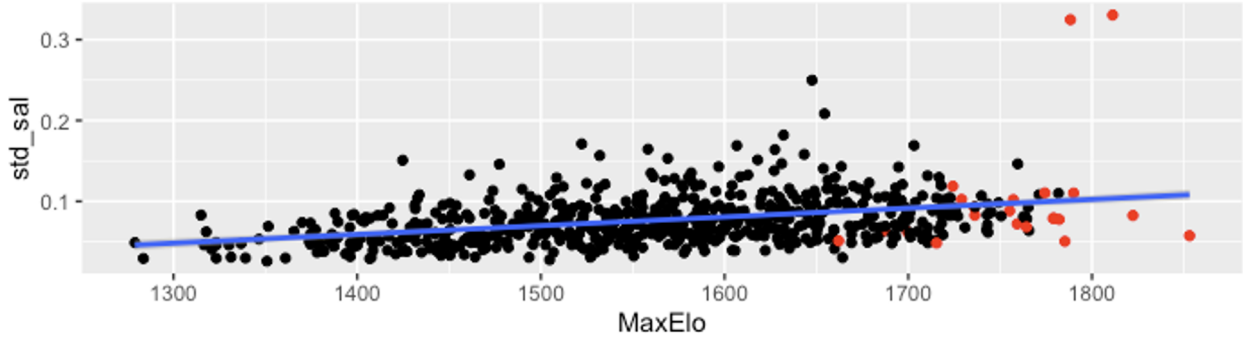
VORP vs Max Elo

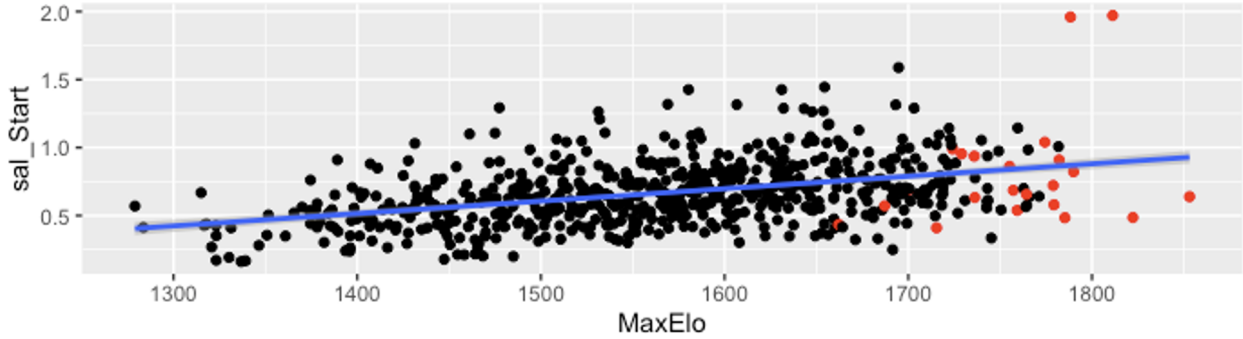




Pure Salary vs Max Elo







Agenda

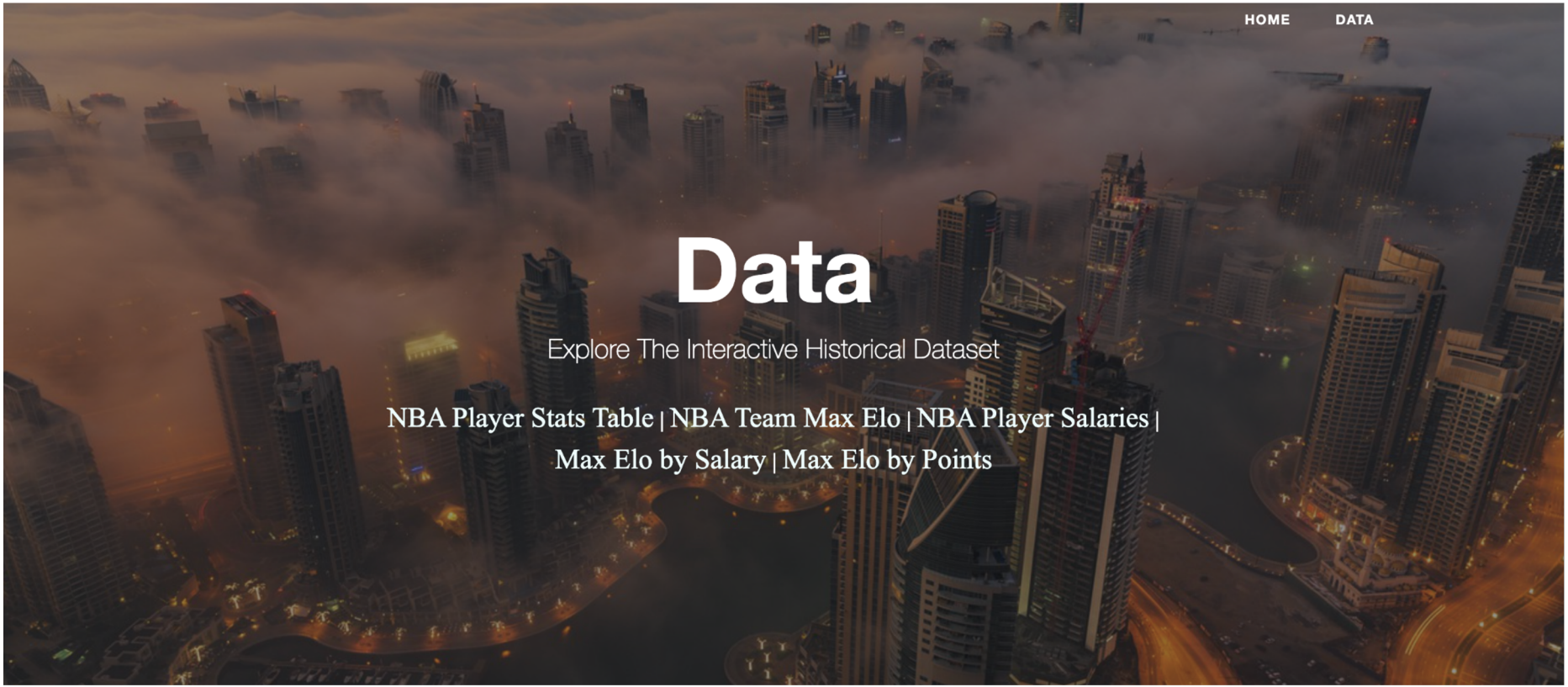
Project Overview
Overview of Project and data sources

Modeling and Variable Selection
Models Used to Predict NBA Team Success

Conclusion
Findings From Analysis and Model

Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation


HOME
DATA
Data
Explore The Interactive Historical Dataset
NBA Player Stats Table | NBA Team Max Elo | NBA Player Salaries |
Max Elo by Salary | Max Elo by Points

NBA Player Stats Table

Player Stats

Player	Tm	2P	2P%	3P	3P%	AST
Alec Burks	UTA	74	43.00%	25	32.90%	30
Boris Diaw	UTA	126	51.20%	20	24.70%	170
Dante Exum	UTA	111	51.90%	44	29.50%	112
Derrick Fav..	UTA	200	49.10%	3	30.00%	56
George Hill	UTA	195	52.30%	94	40.30%	203
Gordon Hay..	UTA	396	50.60%	149	39.80%	252
Jeff Withey	UTA	55	53.90%	0	0.00%	7
Joe Ingles	UTA	81	47.10%	123	44.10%	224
Joe Johnson	UTA	167	45.40%	106	41.10%	144
Joel Bolom..	UTA	8	66.70%	1	25.00%	2
Raul Neto	UTA	31	51.70%	10	32.30%	34
Rodney	UTA	158	44.00%	114	37.10%	96
Rudy Gobert	UTA	413	66.30%	0	0.00%	97
Shelvin	UTA	133	51.00%	37	30.80%	154
Trey Lyles	UTA	94	40.00%	65	31.90%	70

Player

- [ ] Alvin Heggs
- [ ] Alvin Jones
- [ ] Alvin Robertson
- [ ] Alvin Sims
- [ ] Alvin Williams
- [ ] Amal McCaskill
- [ ] Amar'e Stoudemire

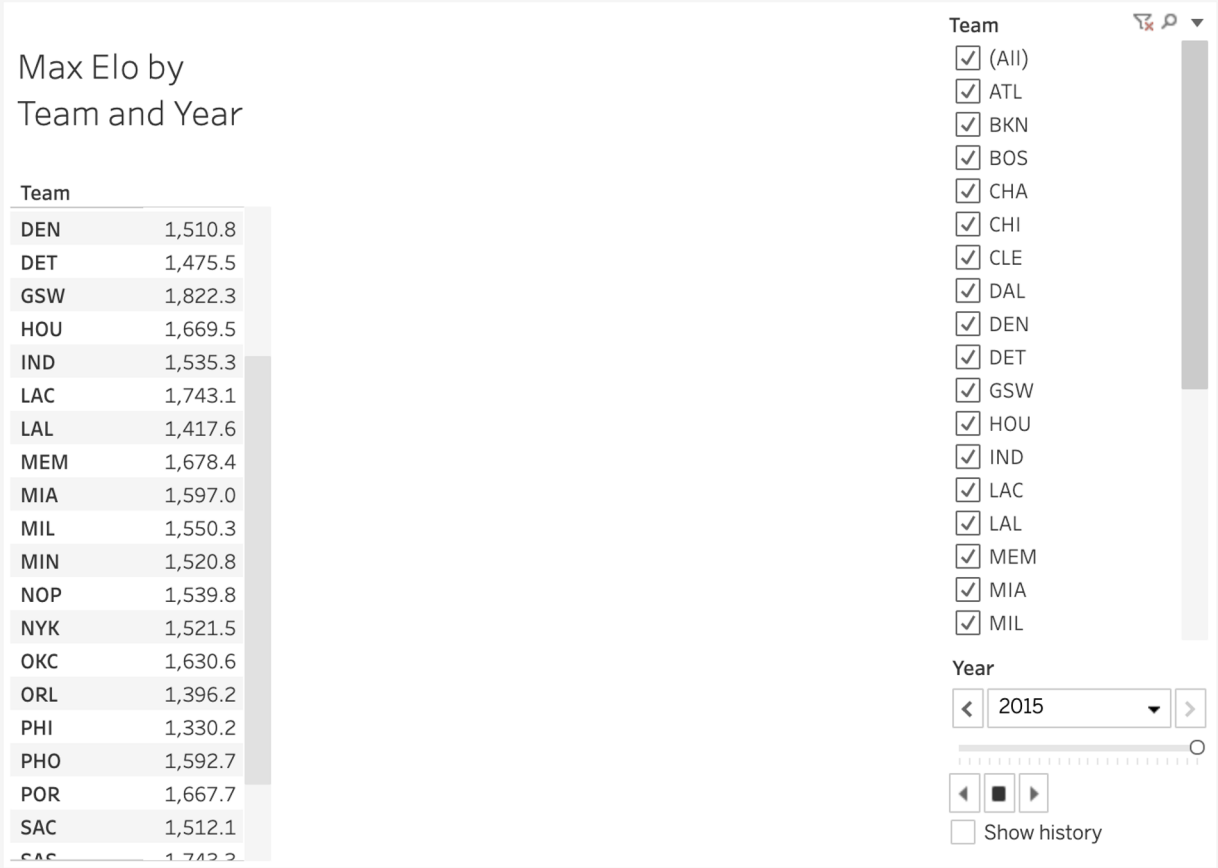
Tm

- [ ] SEA
- [ ] TOR
- [ ] TOT
- [x] UTA
- [ ] VAN
- [ ] WAS
- [ ] WSB

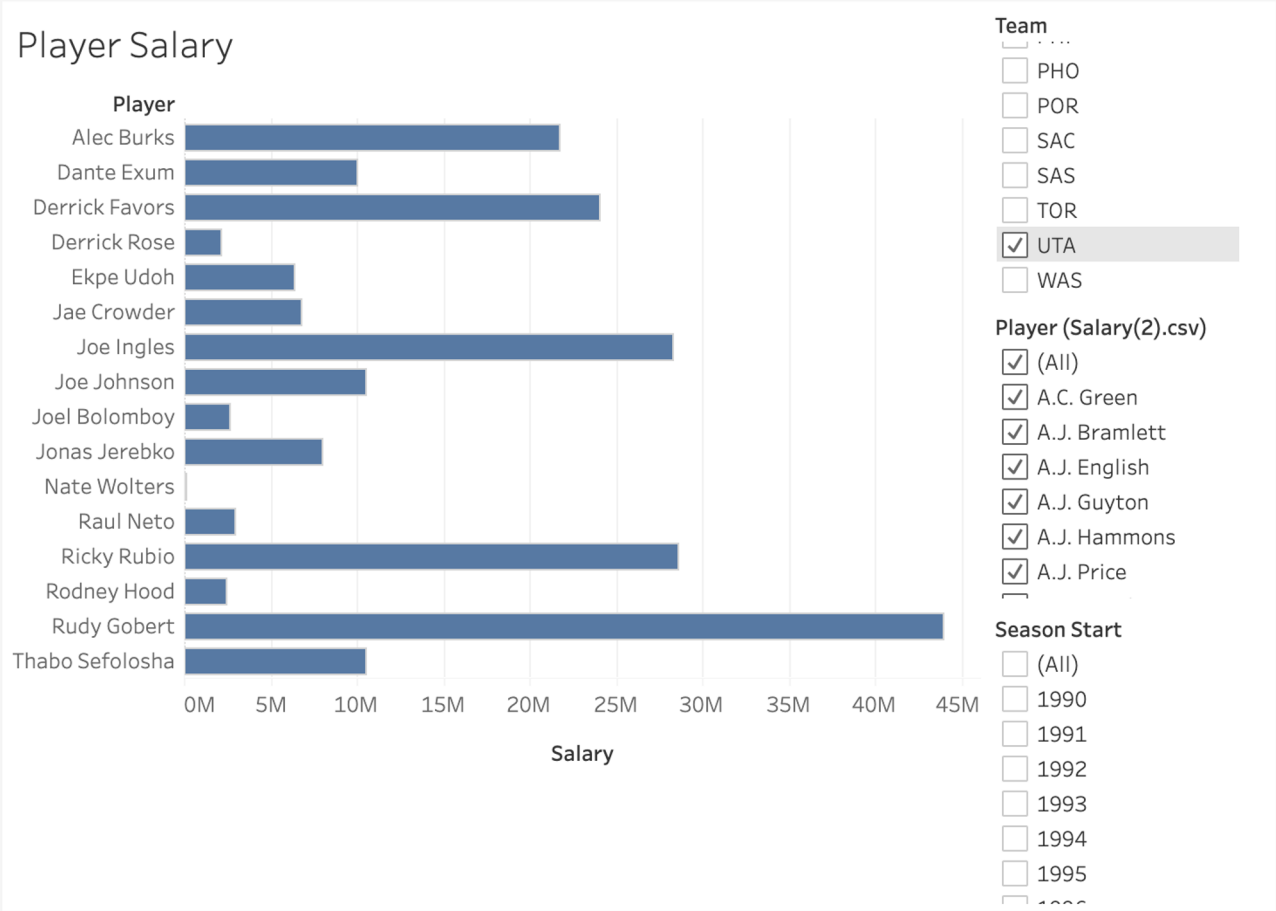
Year

- [ ] 2011
- [ ] 2012
- [ ] 2013
- [ ] 2014
- [ ] 2015
- [ ] 2016
- [x] 2017

NBA Team Max Elo



NBA Player Salaries

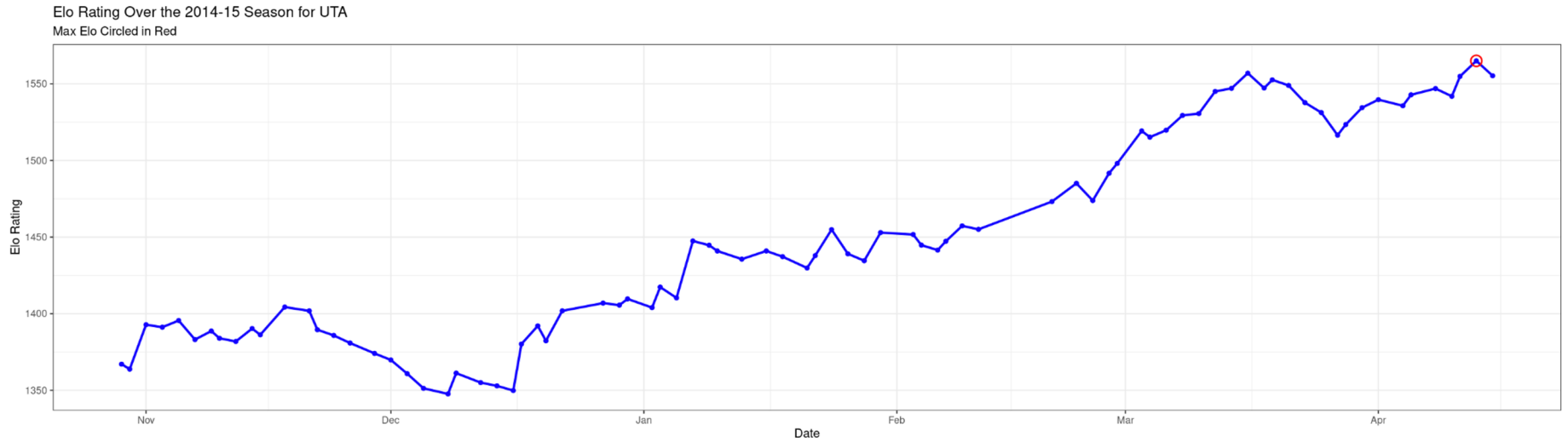


Predicting Max Elo Rating using NBA Salaries

Select Team

UTA-2015

Name	Total_Salary_Cap	Total_Cap_per_VORP	Predicted_Max_Elo_Rating	Actual_Max_Elo_Rating	In_Range	Predicted_Category	Actual_Category
UTA-2015	75.99%	561.79%	1542.9	1565	YES	Average Team	Playoff Bound



Agenda

Project Overview
Overview of Project and data sources

Modeling and Variable Selection
Models Used to Predict NBA Team Success

Conclusion
Findings From Analysis and Model

Website Walkthrough
Data deep-dive

Further Exploration
Potential improvements and other areas of investigation

Further Exploration

- Player Injury
- CBA Rules
- Changing Rules over Time
- Starters vs Bench Players

Questions