NBA Predictive Model

By Matt Pucci

Data and Methodology

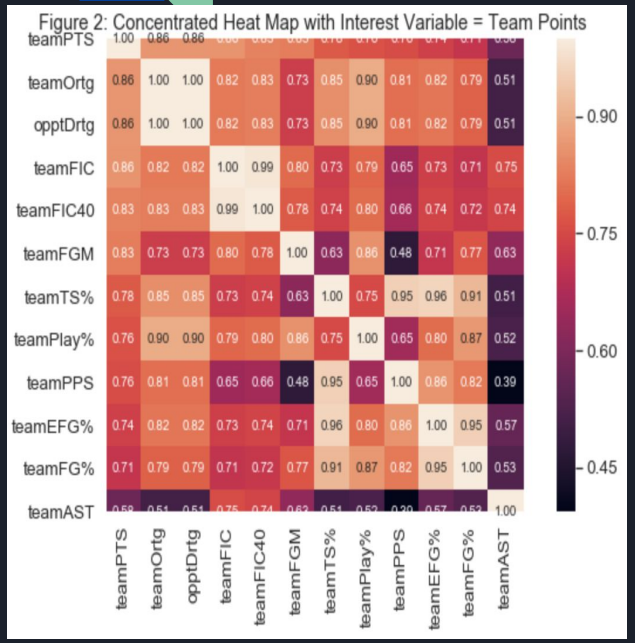
Train Data Head:

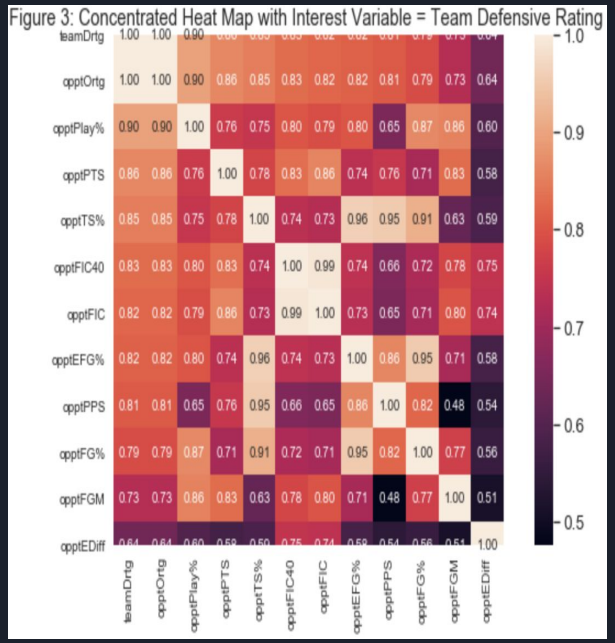
	gmDate	gmTime	seasTyp	offLNm1	offFNm1	offLNm2	offFNm2	offLNm3	offFNm3	teamAbbr	...	opptFIC40	opptOrtg	opptDrtg	opptEDiff	opptPlay%
0	2016-10-25	08:00	Regular	Lane	Karl	Adams	Bennie	Kennedy	Bill	NY	...	84.9585	116.5653	87.6731	28.8922	0.4592
1	2016-10-25	08:00	Regular	Lane	Karl	Adams	Bennie	Kennedy	Bill	CLE	...	41.6667	87.6731	116.5653	-28.8922	0.3478
2	2016-10-25	10:00	Regular	Buchert	Nick	Callahan	Mike	Brown	Tony	UTA	...	69.6653	122.1120	112.3863	9.7257	0.4699
3	2016-10-25	10:00	Regular	Buchert	Nick	Callahan	Mike	Brown	Tony	POR	...	58.7137	112.3863	122.1120	-9.7257	0.4444
4	2016-10-25	10:30	Regular	Maddox	Tre	Fraher	Pat	Crawford	Dan	SA	...	60.7884	101.2387	130.5979	-29.3592	0.4301

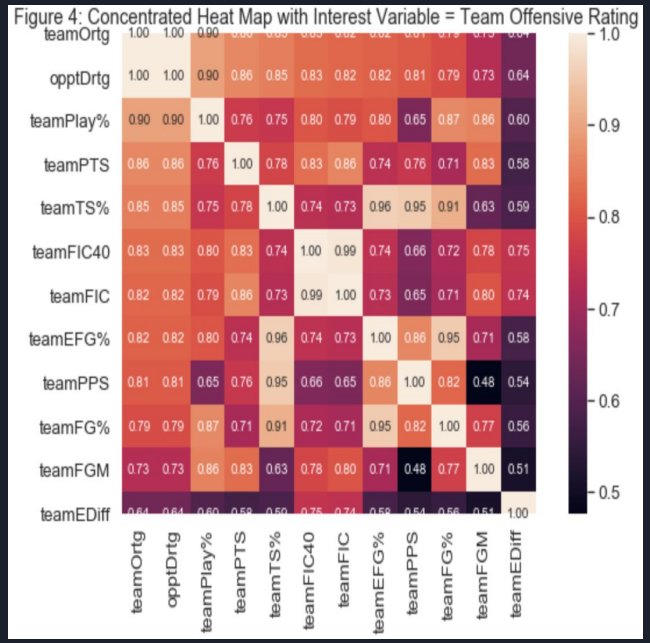
Test Data Head:

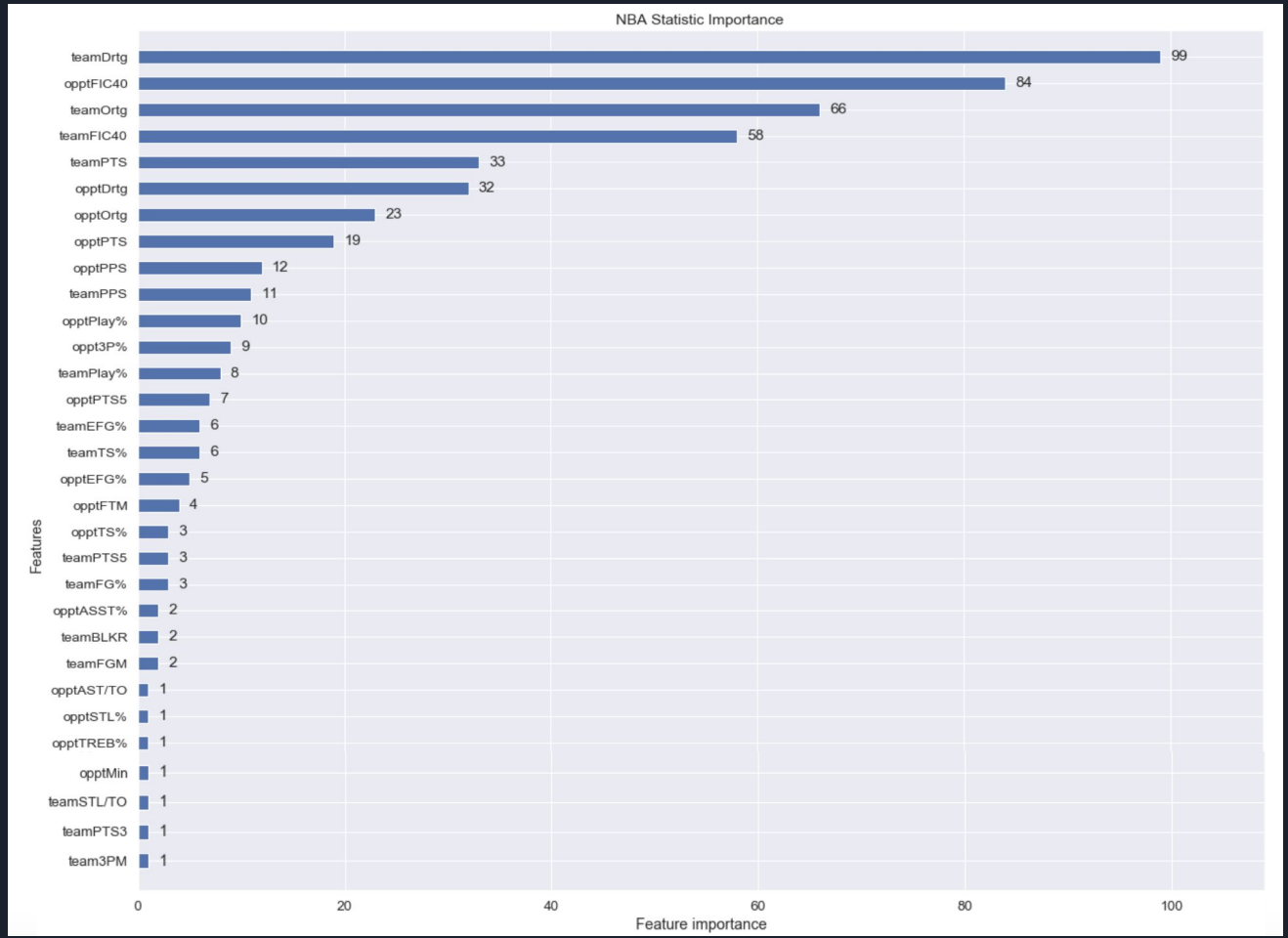
gmDate	gmTime	seasTyp	offLNm1	offFNm1	offLNm2	offFNm2	offLNm3	offFNm3	teamAbbr	...	opptFIC40	opptOrtg	opptDrtg	opptEDiff	opptPlay%
2017-10-17	08:00	Regular	Forte	Brian	Smith	Michael	McCutchen	Monty	BOS	...	55.5208	101.7143	98.7227	2.9916	0.4176
2017-10-17	08:00	Regular	Forte	Brian	Smith	Michael	McCutchen	Monty	CLE	...	62.5519	98.7227	101.7143	-2.9916	0.3956
2017-10-17	10:30	Regular	Maddox	Tre	Garretson	Ron	Foster	Scott	HOU	...	86.2033	118.0353	119.0108	-0.9755	0.4725
2017-10-17	10:30	Regular	Maddox	Tre	Garretson	Ron	Foster	Scott	GS	...	81.9038	119.0108	118.0353	0.9755	0.4700
2017-10-18	07:00	Regular	Davis	Marc	Boland	Matt	DeRosa	Joe	CHA	...	70.7113	103.0506	90.9270	12.1236	0.4271

Data Exploration: Seaborn HeatMaps








NBA Statistic Importance
teamDrtg
opptFIC40
teamOrtg
teamFIC40
teamPTS
opptDrtg
opptOrtg
opptPTS
opptPPS
teamPPS
opptPlay%
oppt3P%
teamPlay%
opptPTS5
teamEFG%
teamTS%
opptEFG%
opptFTM
opptTS%
teamPTS5
teamFG%
opptASST%
teamBLKR
teamFGM
opptAST/TO
opptSTL%
opptTREB%
opptMin
teamSTL/TO
teamPTS3
team3PM
99
84
66
58
33
32
23
19
12
11
10
9
8
7
6
6
5
4
3
3
3
2
2
2
1
1
1
1
1
1
1
Features
0
20
40
60
80
100
Feature importance

Model Performance

Model	Performance
KNN	0.8130081300813008
Random Forest Classifier	0.8211382113821138
Gradient Boost Classifier	0.8414634146341463
Gradient Boost Classifier (x_new, y_new)	0.866260162601626

Model Results

```
CLE vs BOS
['Win']
[[0.34947394 0.65052606]]
------------------------------

HOU vs GS
['Loss']
[[0.78913164 0.21086836]]
------------------------------

DET vs CHA
['Win']
[[0.03788526 0.96211474]]
------------------------------

IND vs BKN
['Win']
[[0.05419061 0.94580939]]
------------------------------

ORL vs MIA
['Win']
[[0.13727335 0.86272665]]
------------------------------
```

```
WAS vs PHI
['Win']
[[0.35992836 0.64007164]]
------------------------------

BOS vs MIL
['Win']
[[0.44835512 0.55164488]]
------------------------------

MEM vs NO
['Win']
[[0.38951225 0.61048775]]
------------------------------

DAL vs ATL
['Loss']
[[0.92380546 0.07619454]]
------------------------------

UTA vs DEN
['Loss']
[[0.56538266 0.43461734]]
------------------------------
```

Future Work

- Clearly this model is still in its early stages and it does not capture all of the factors that go into winning a NBA game.
- Some features I would like to improve on:
  - Importance of game - rivalry, personal inspiration for players.
  - Injuries - As mentioned the model fails to be able to capture any injuries that would severely affect a team.
  - Referees - It is well known that some players have issues with certain referees. I would like to be able to quantify this and apply it to the model.
- Spread Calculator - Although this model is in a good starting position, I would like to add a feature that predicts the point spread so it can truly be used as a Sports Bettor tool.