Explanation document:

Before I've started to analyze I had to organize all the results in an excel file (Attached).

First step:

Data Explanation:

1. Choosing classes- the target in this project is to calculate the probability the Warriors will win the game. The classes are divided in to two categories: win and loss. I've checked the labels quantity, so we can decide the score for the model– the data is balanced (figure 1), so I'll maximize accuracy.

2. Estimate the features - I noticed that every feature has a float value except the name of the team (feature game), so I dropped this feature.
Also, I decided to ignore the features 'FG Made-Attempted 1', 'FG Made-Attempted 2', '3PT Made-Attempted 1', '3PT Made-Attempted 2', 'FT Made-Attempted 1', 'FT Made-Attempted 2', since the features are pointing the total points for each team - which will tell us who is the winning team.
For each feature I did a statistic description and a box plot - I saw that the feature 'Flagrant Fouls 1' has no impact because this column is a zero column, and I've also succeeded to identify outliers.
The feature 'win' will be the target of the model, so I've separated it from the features.

3. Data Distribution- I performed tests of the various data distributions (for each feature) using graphs (KDE plots) that help us to understand the behavior of the data relative to each category. These distributions can help us understand which features will be most effective in distinguishing between the two classes (win and loss) - the features in which the greatest difference is between the distributions. In this case, the distributions are relatively similar. We also can see that there is

4. Correlation matrix- In this section I examined whether there is a correlation between the various features. A high correlation indicates that we should consider removing features, as they do not explain new information. After the test we found that the parameters are not correlated to each other, the correlations found to be less than 0.4.

second step:

building a decision tree:

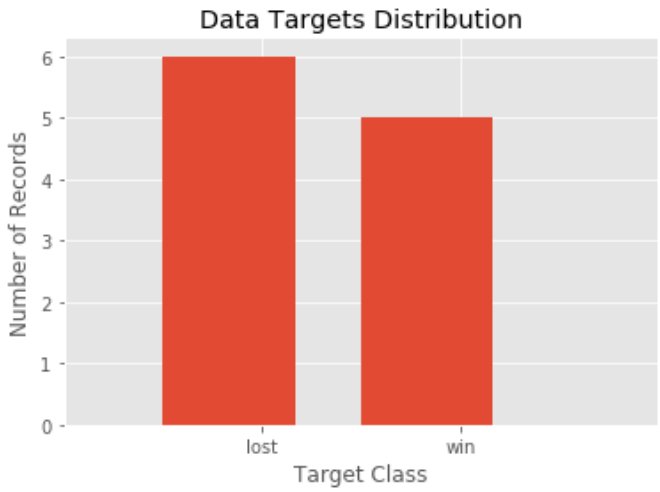
I performed a decision tree in order to see what features the tree would choose to be meaningful for separating the win.

Third step:

Model:
the model I decided is Random Forest. I think it's most suitable, because this model is using ensembles of trees, where each tree in the ensemble is grown in accordance with a random feature. Final predictions are obtained by aggregating over the ensemble. As the base constituents of the ensemble are tree-structured predictors, and since each of these trees is constructed using an injection of randomness. RF model have succeeded to predict the target in many domains, so with more data it can do an excelnt work in the project domain.
For building the model we'll look at the important variables that contribute to the win. After we have a look at the important predictor variables, we need to check the win probability predictions for the final match.

I read that Artificial Neural Network is also suitable for this project, because I didn't learn about this model I didn't want to choose this model, but from what I read this model.

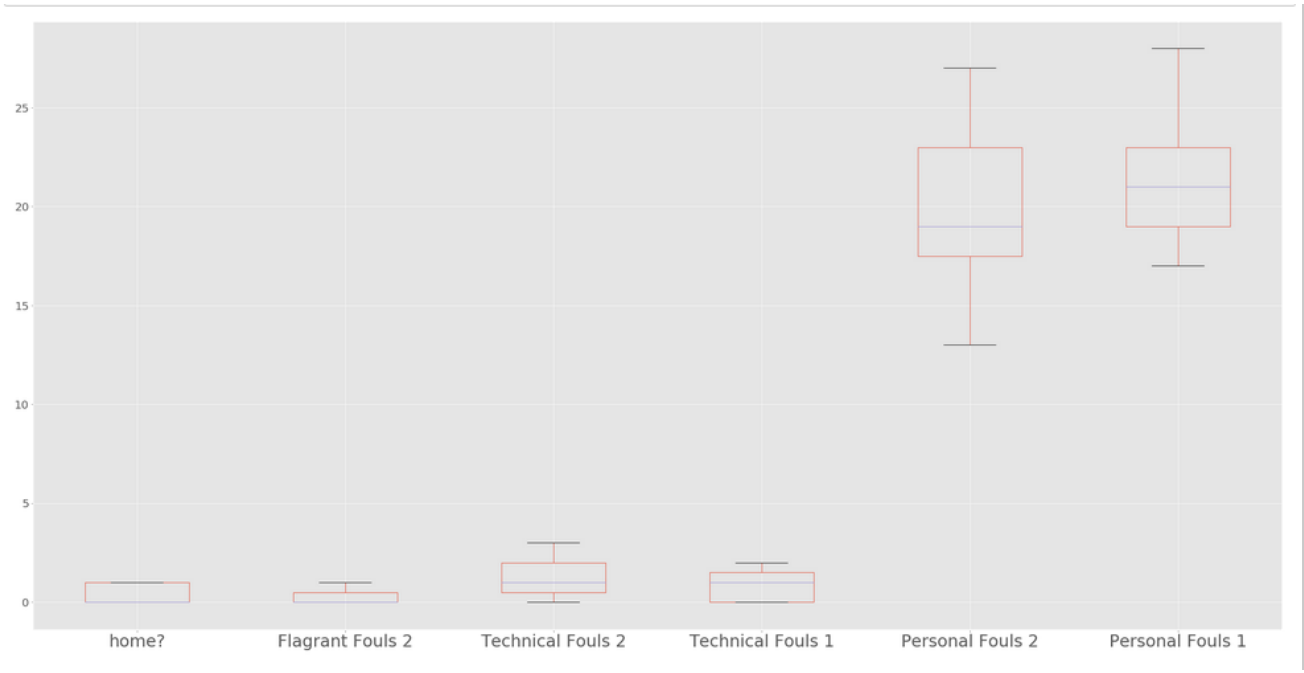
attachments photos:

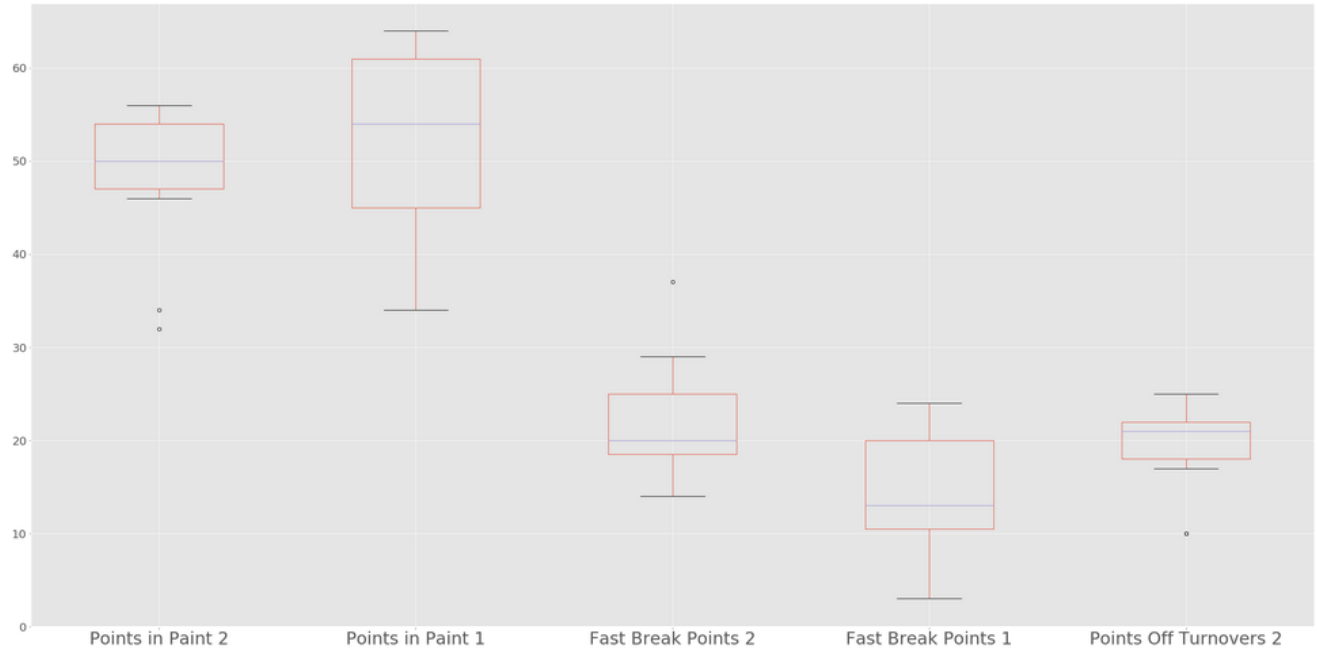


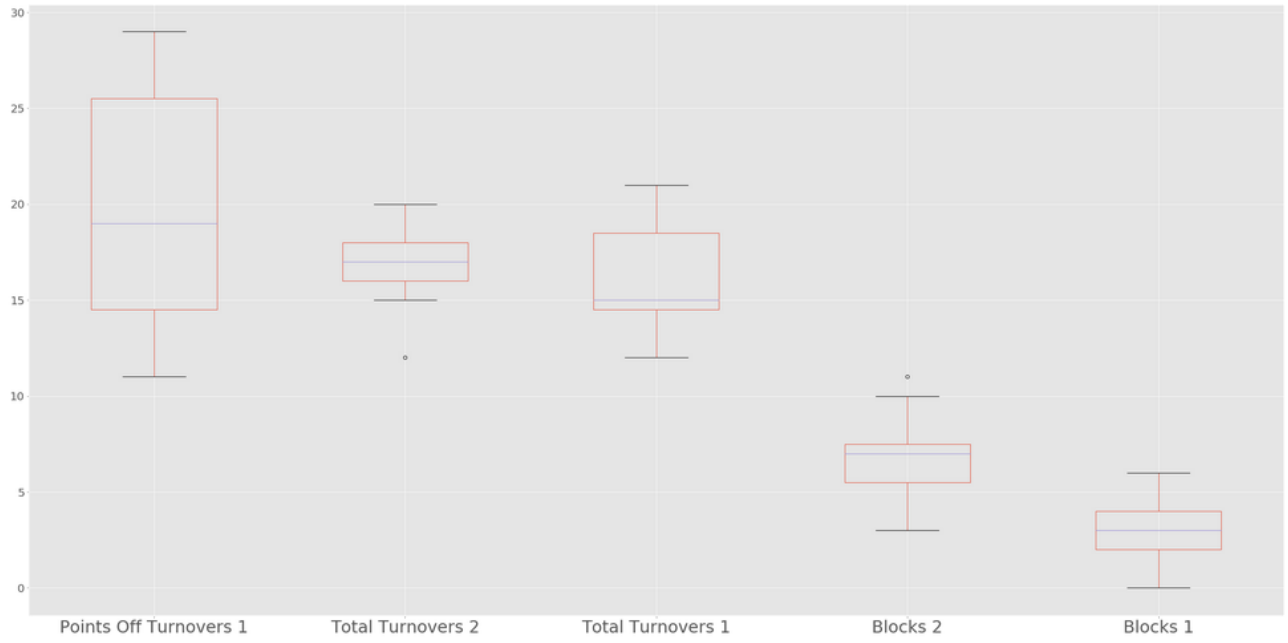
We can declare that the classes are well-balanced.

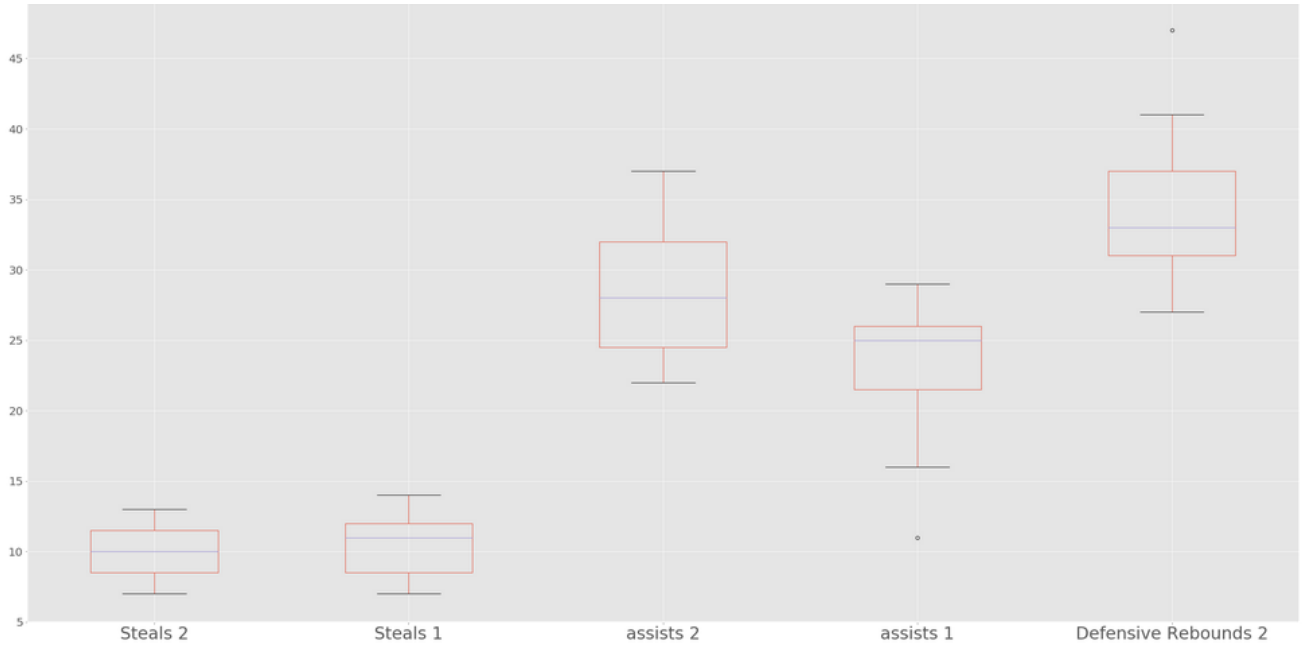
Description and Box Plot:

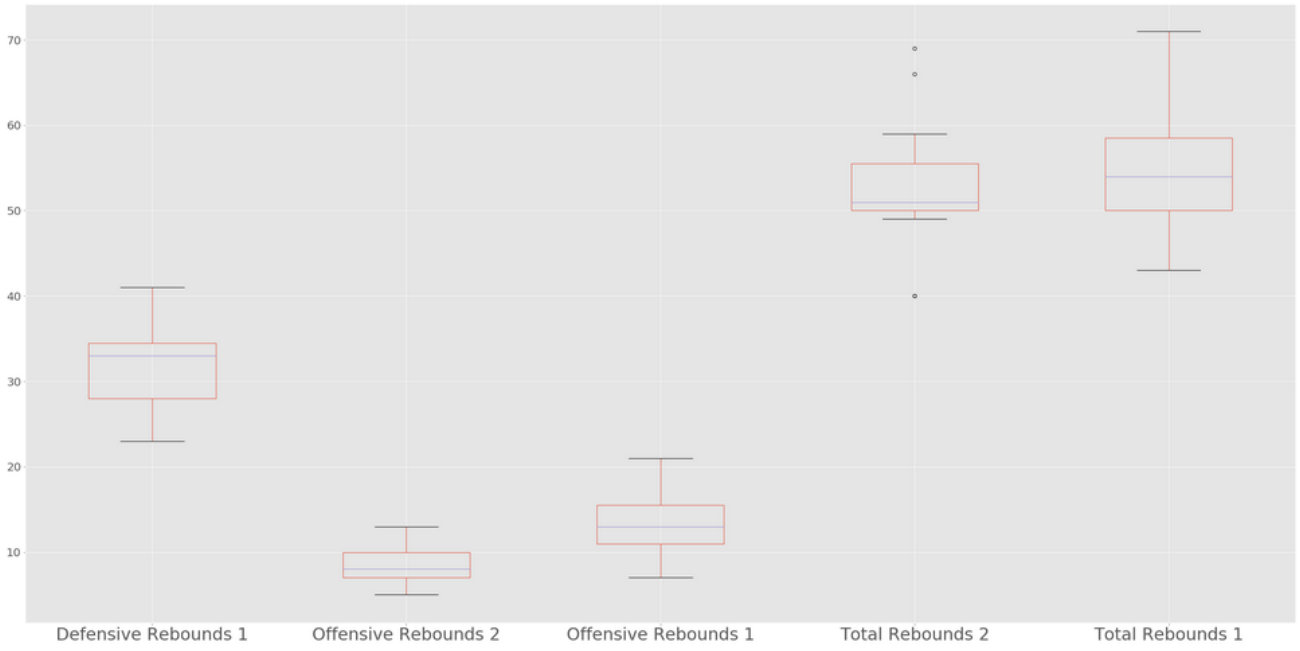
	home?	win?	Flagrant Fouls 2	Flagrant Fouls 1	Technical Fouls 2	Technical Fouls 1	Personal Fouls 2	Personal Fouls 1	Points in Paint 2	Points in Paint 1	Fast Break Points 2	Fast Break Points 1	Points Off Turnovers 2	F T
count	11.000000	11.000000	11.000000	11.0	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	11.000000	1
mean	0.454545	0.454545	0.272727	0.0	1.181818	0.909091	20.000000	21.545455	47.818182	52.181818	22.181818	14.363636	19.181818	1
std	0.522233	0.522233	0.467099	0.0	0.981650	0.831209	4.219005	3.777926	7.972681	10.215852	6.554665	7.201010	4.996362	
min	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	13.000000	17.000000	32.000000	34.000000	14.000000	3.000000	10.000000	1
25%	0.000000	0.000000	0.000000	0.0	0.500000	0.000000	17.500000	19.000000	47.000000	45.000000	18.500000	10.500000	18.000000	1
50%	0.000000	0.000000	0.000000	0.0	1.000000	1.000000	19.000000	21.000000	50.000000	54.000000	20.000000	13.000000	21.000000	1
75%	1.000000	1.000000	0.500000	0.0	2.000000	1.500000	23.000000	23.000000	54.000000	61.000000	25.000000	20.000000	22.000000	2
max	1.000000	1.000000	1.000000	0.0	3.000000	2.000000	27.000000	28.000000	56.000000	64.000000	37.000000	24.000000	25.000000	2

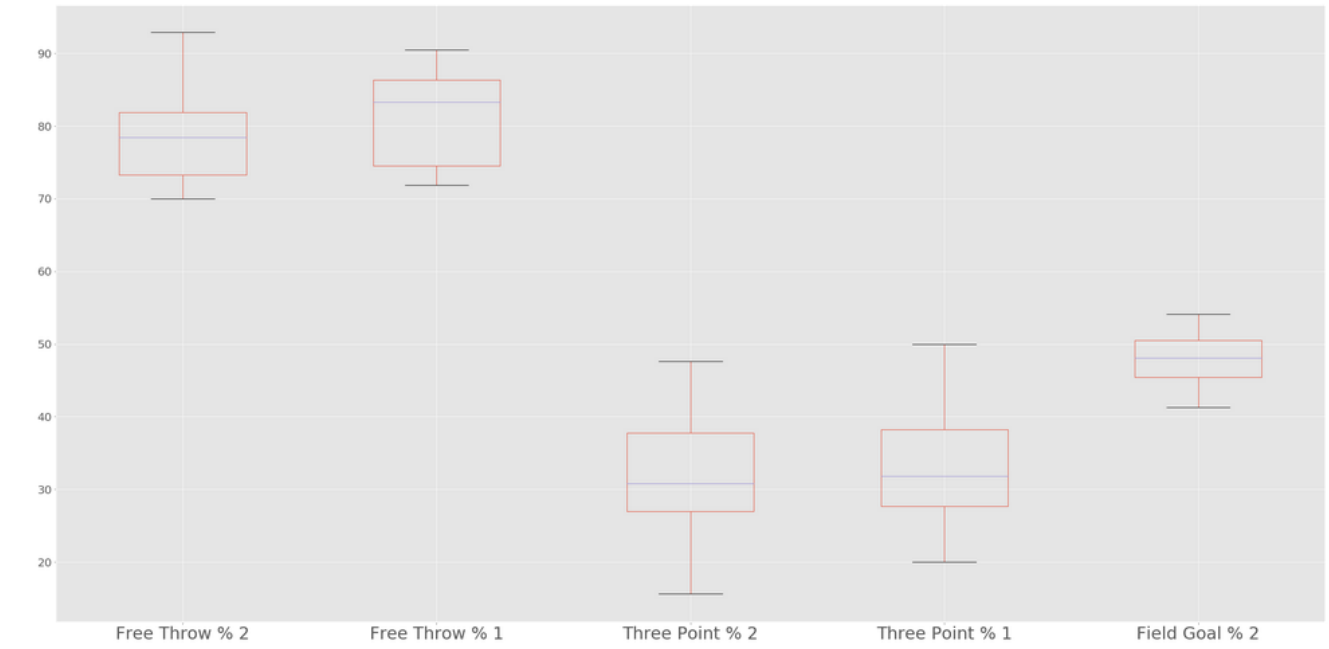

25
20
15
10
5
0
home?
Flagrant Fouls 2
Technical Fouls 2
Technical Fouls 1
Personal Fouls 2
Personal Fouls 1


60
50
40
30
20
10
0
Points in Paint 2
Points in Paint 1
Fast Break Points 2
Fast Break Points 1
Points Off Turnovers 2

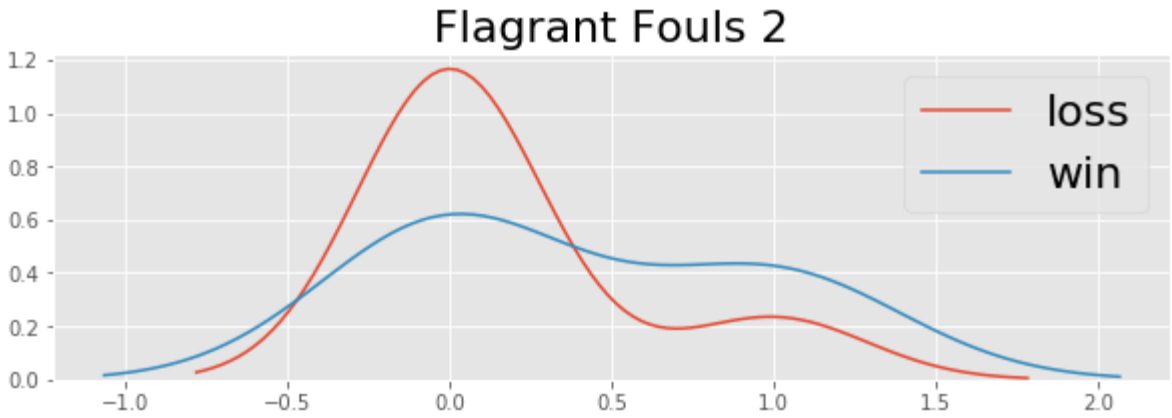

30
25
20
15
10
5
0
Points Off Turnovers 1
Total Turnovers 2
Total Turnovers 1
Blocks 2
Blocks 1

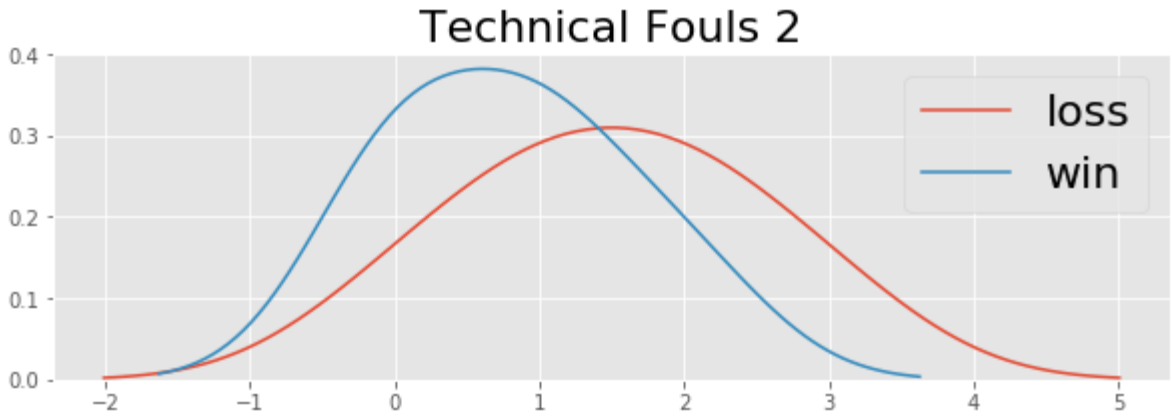

45
40
35
30
25
20
15
10
5
Steals 2
Steals 1
assists 2
assists 1
Defensive Rebounds 2

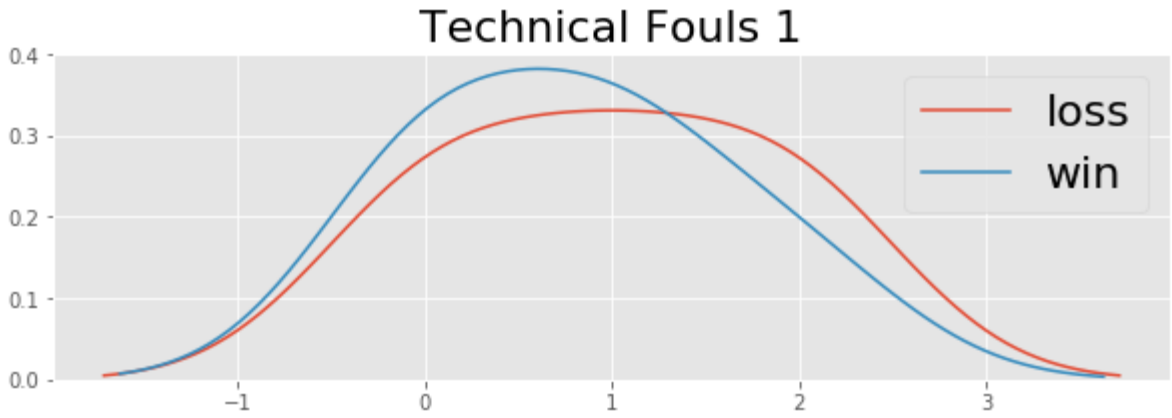

70
60
50
40
30
20
10
Defensive Rebounds 1
Offensive Rebounds 2
Offensive Rebounds 1
Total Rebounds 2
Total Rebounds 1

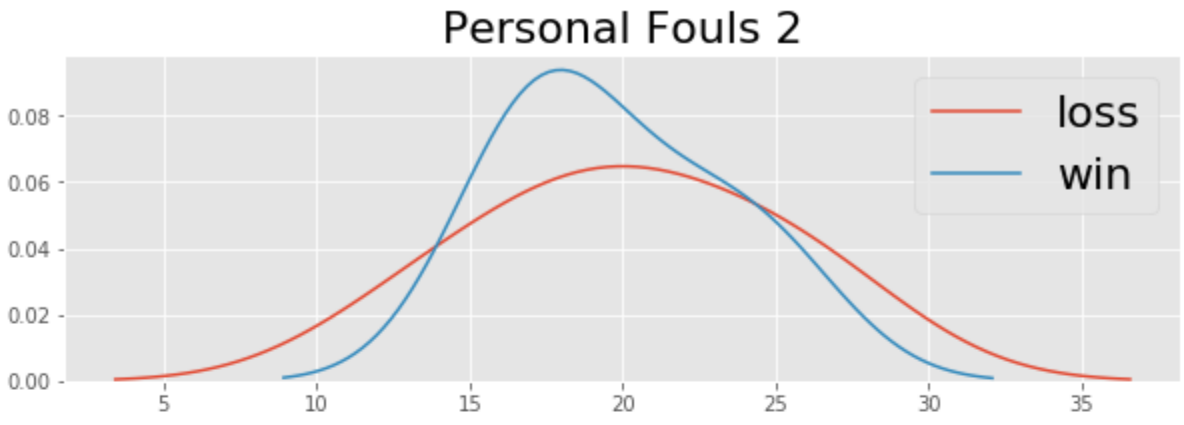

90
80
70
60
50
40
30
20
Free Throw % 2
Free Throw % 1
Three Point % 2
Three Point % 1
Field Goal % 2

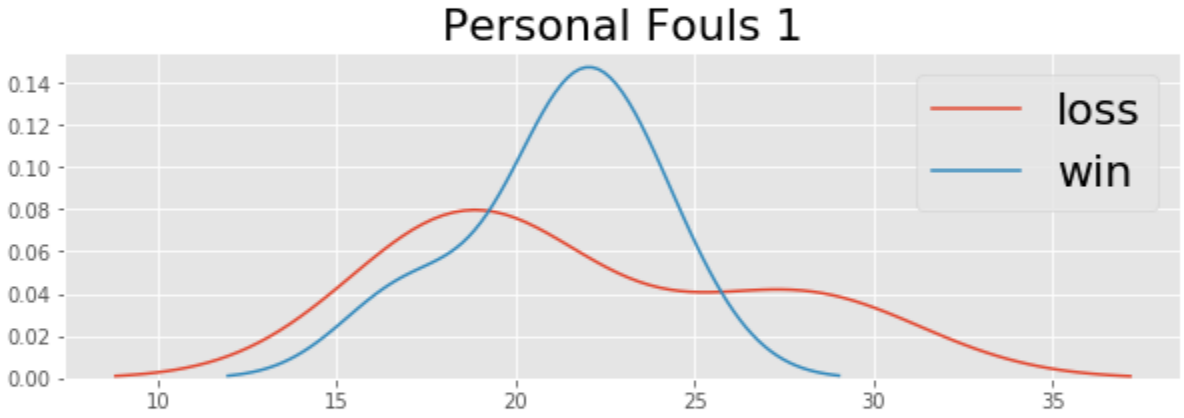
Distribution of each feature within the classes

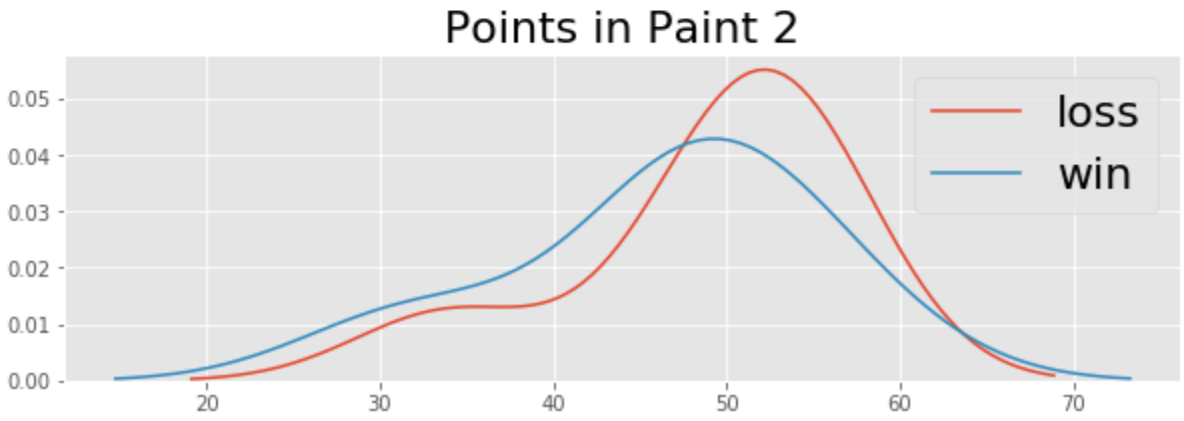


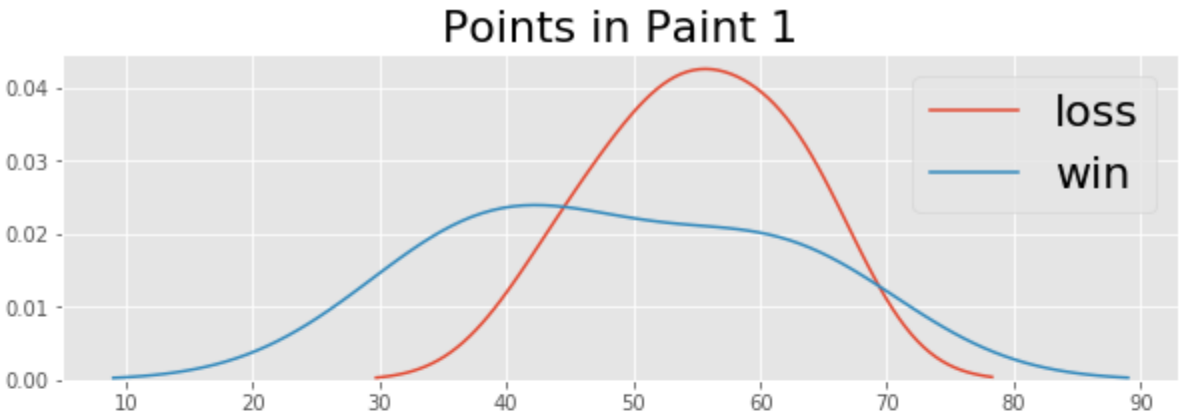


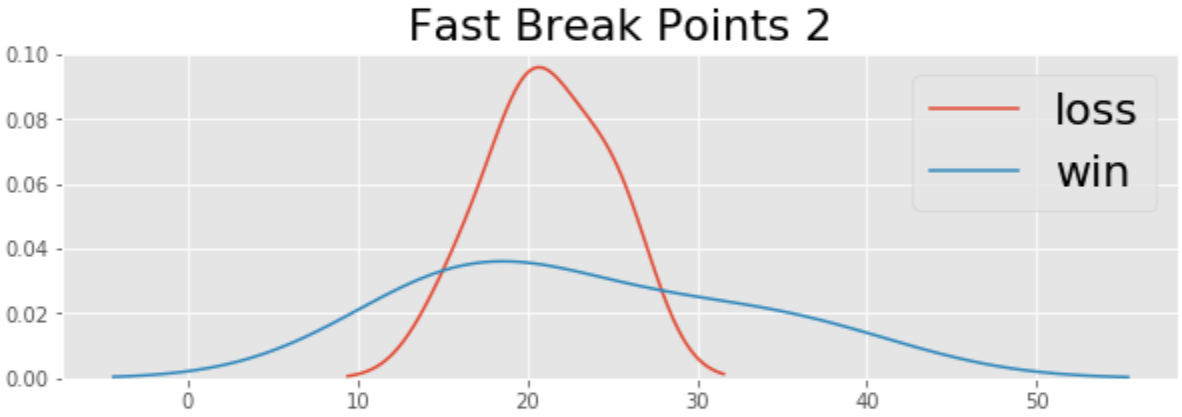


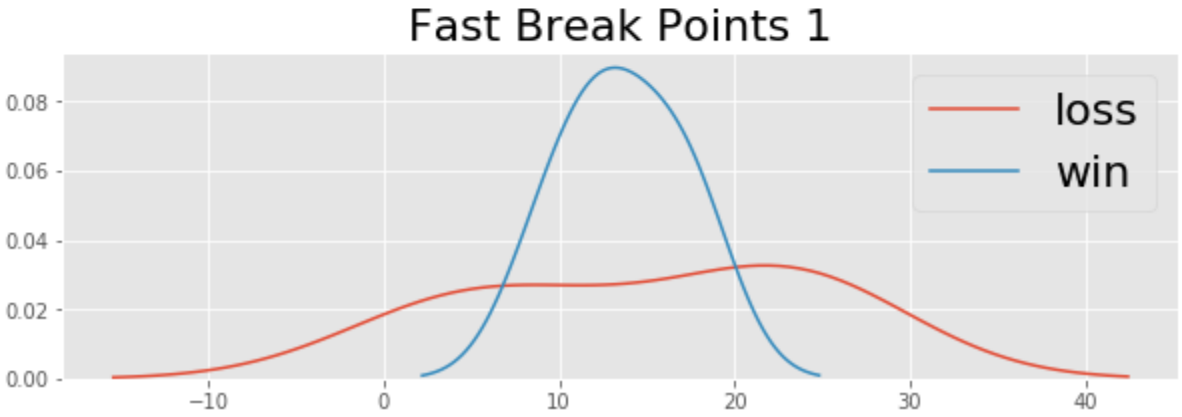

Personal Fouls 2
loss
win
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25
30
35

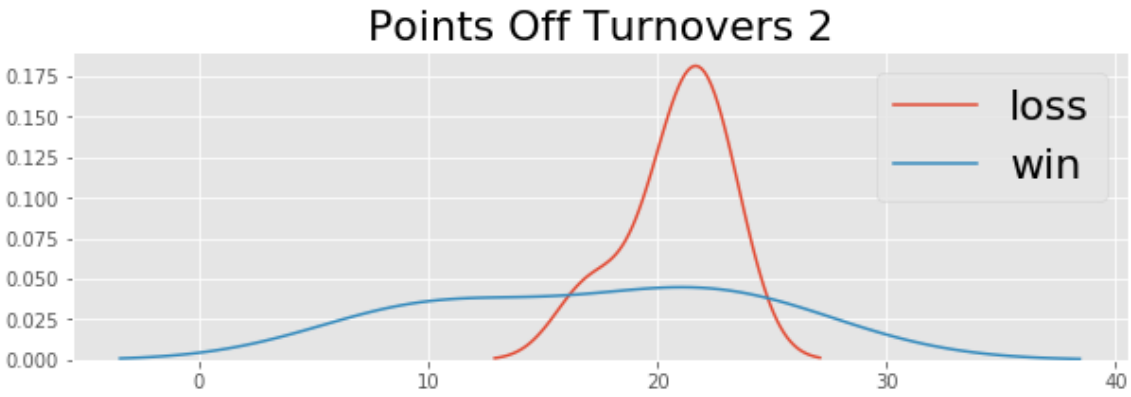

Personal Fouls 1
loss
win
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10
15
20
25
30
35

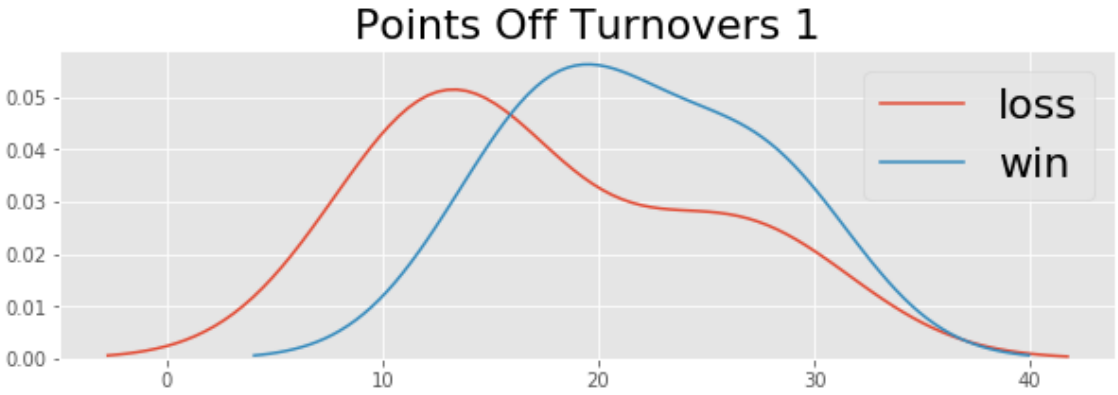

Points in Paint 2
loss
win
0.05
0.04
0.03
0.02
0.01
0.00
20
30
40
50
60
70

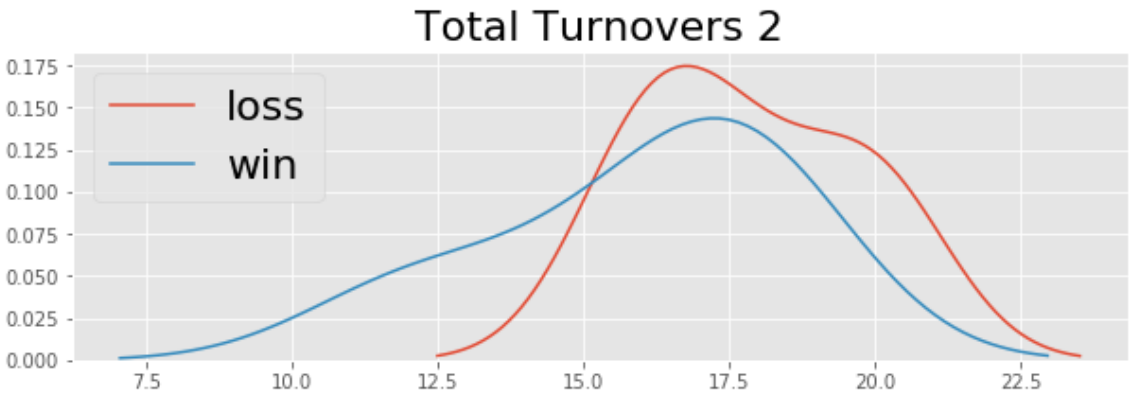

Points in Paint 1
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60
70
80
90
loss
win

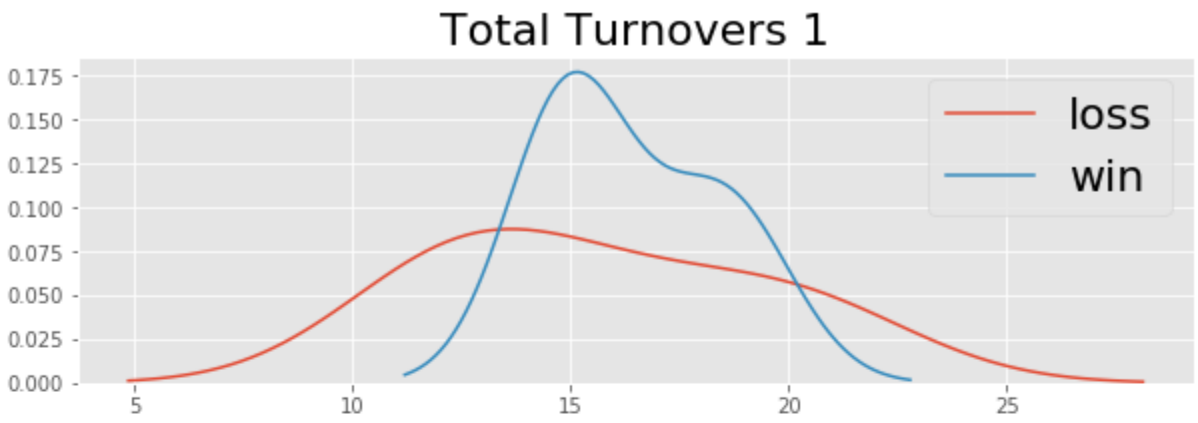

Fast Break Points 2
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
loss
win

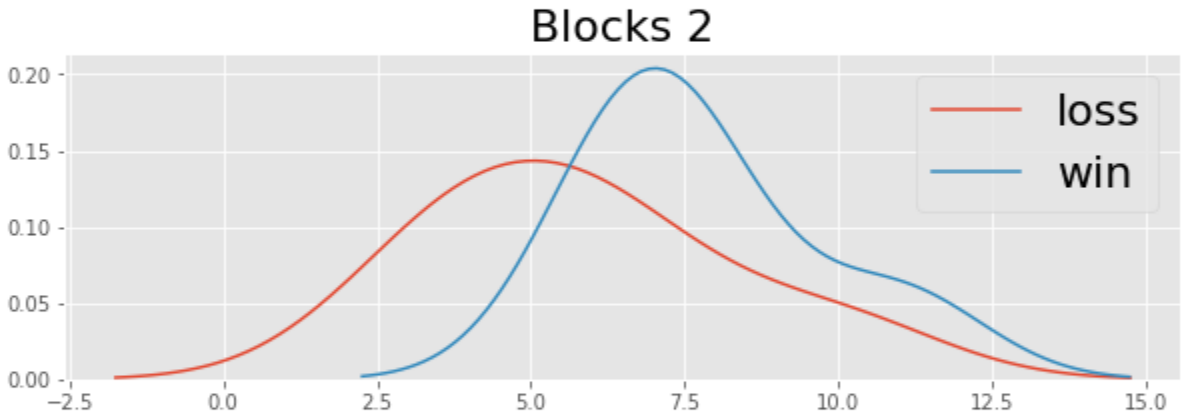

Fast Break Points 1
0.08
0.06
0.04
0.02
0.00
−10
0
10
20
30
40
loss
win

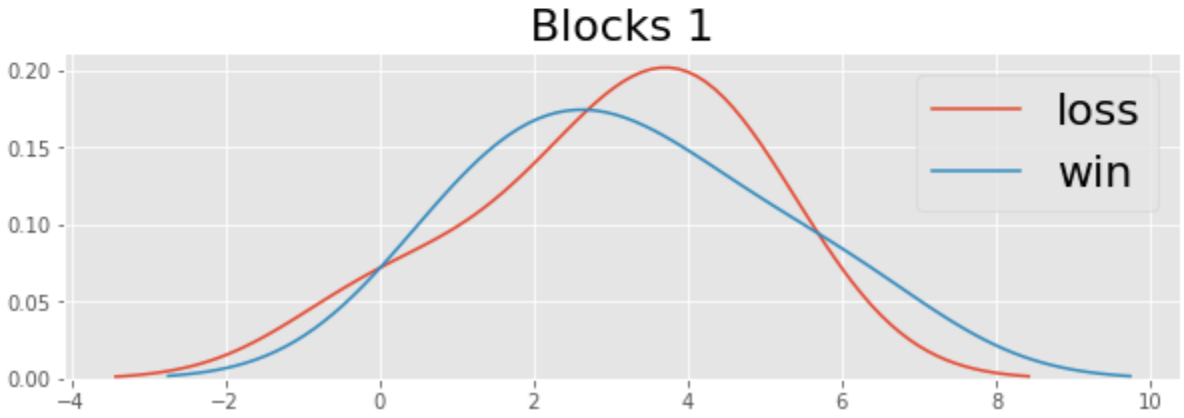

Points Off Turnovers 2
loss
win
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
10
20
30
40

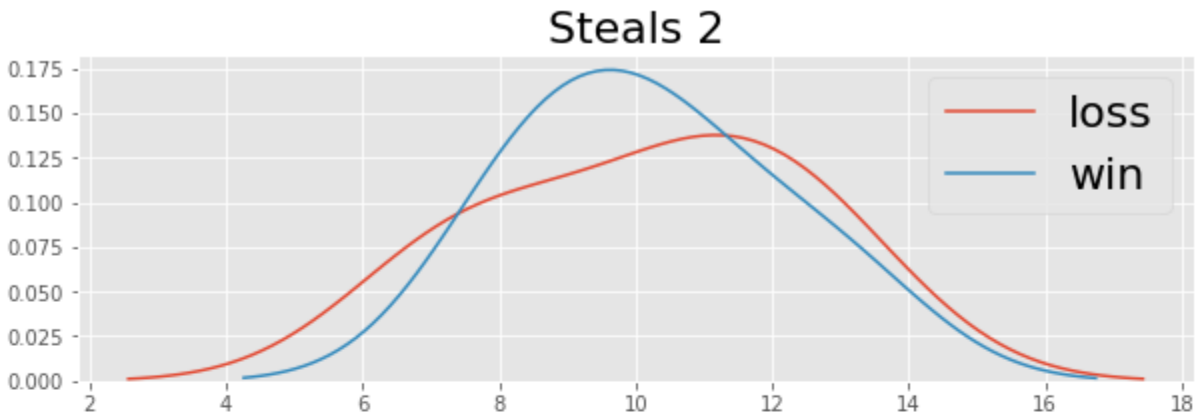

Points Off Turnovers 1
loss
win
0.05
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40

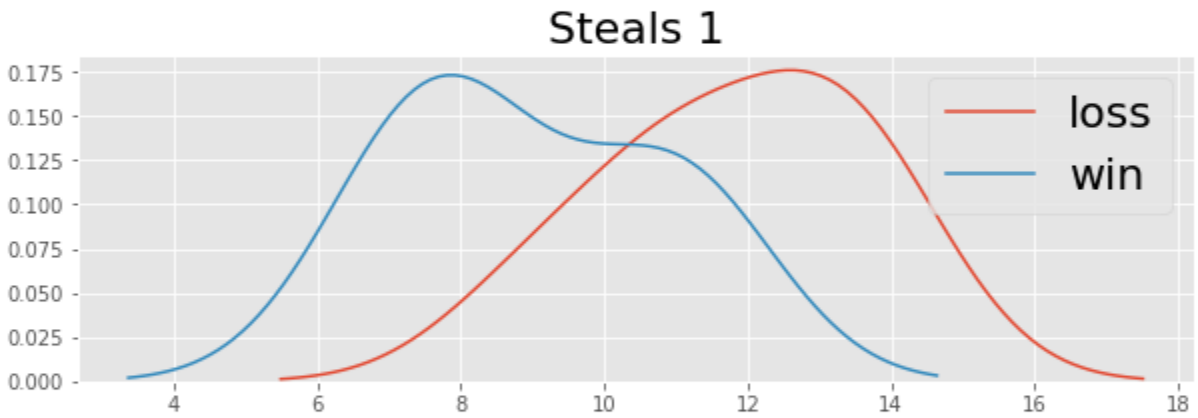

Total Turnovers 2
loss
win
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
7.5
10.0
12.5
15.0
17.5
20.0
22.5

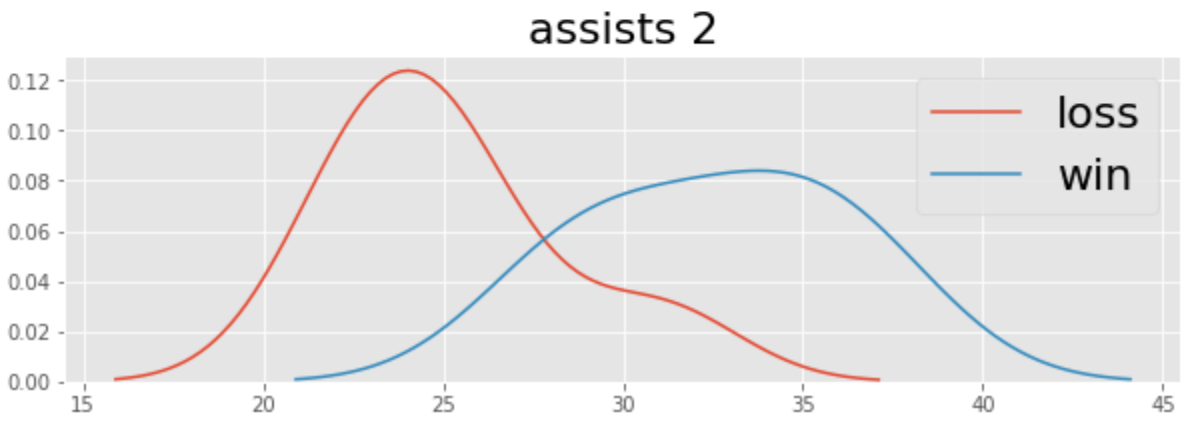

Total Turnovers 1
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
5
10
15
20
25
loss
win

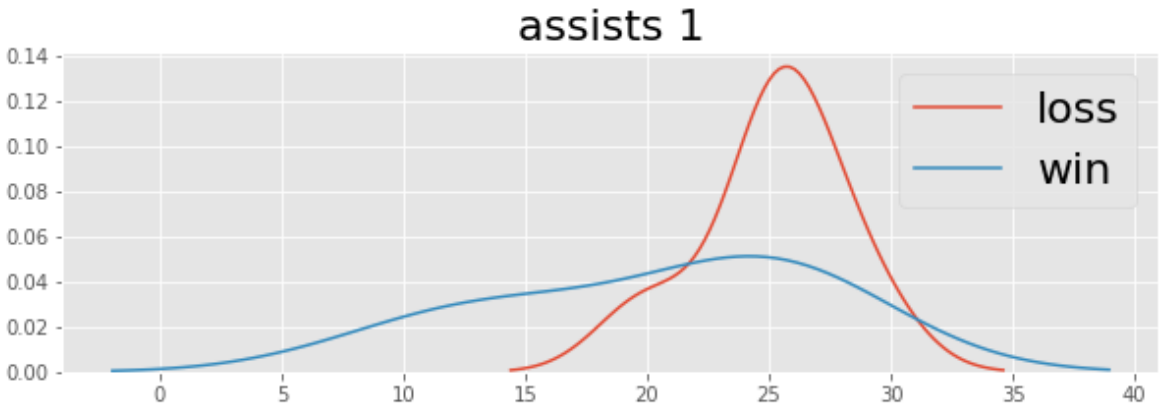

Blocks 2
0.20
0.15
0.10
0.05
0.00
−2.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
loss
win

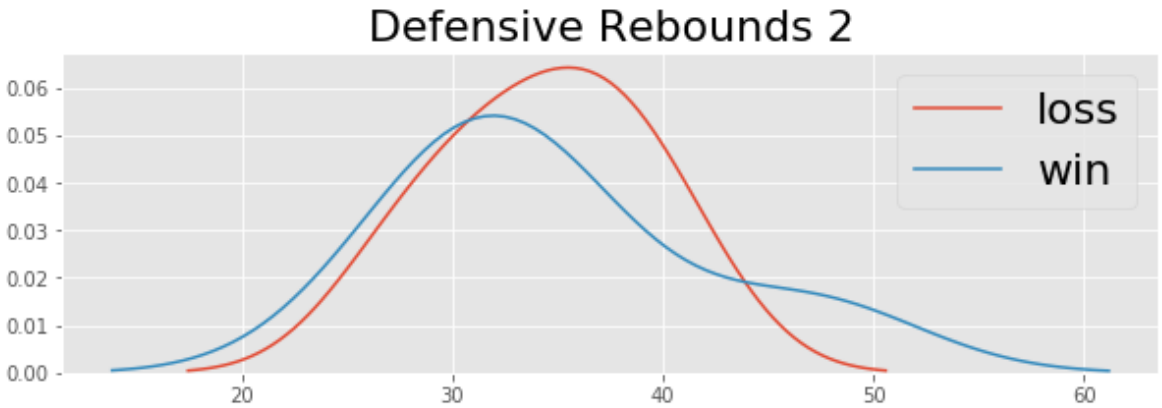

Blocks 1
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4
6
8
10
loss
win

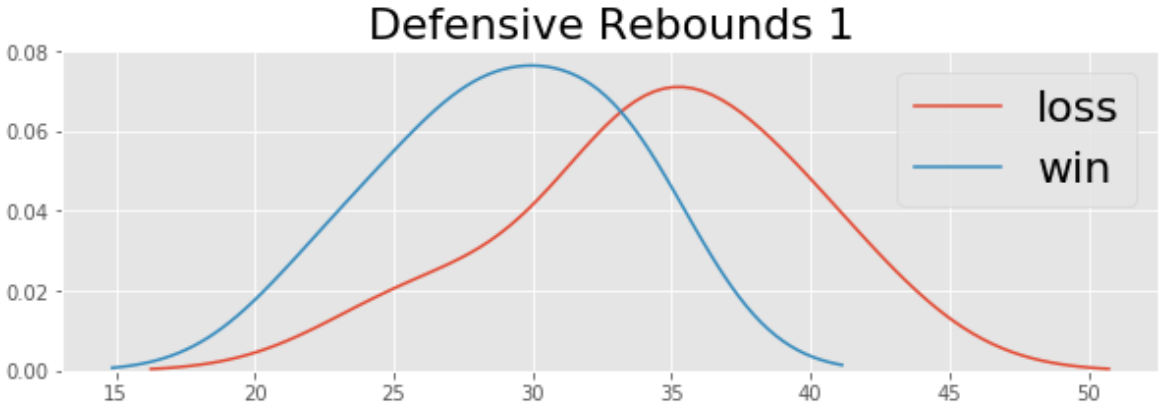

Steals 2
loss
win
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
2
4
6
8
10
12
14
16
18

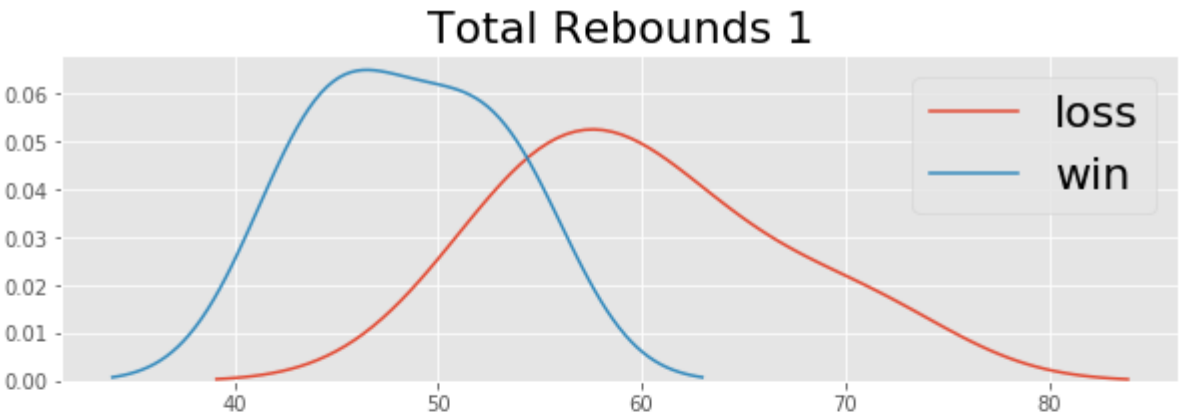

Steals 1
loss
win
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
4
6
8
10
12
14
16
18

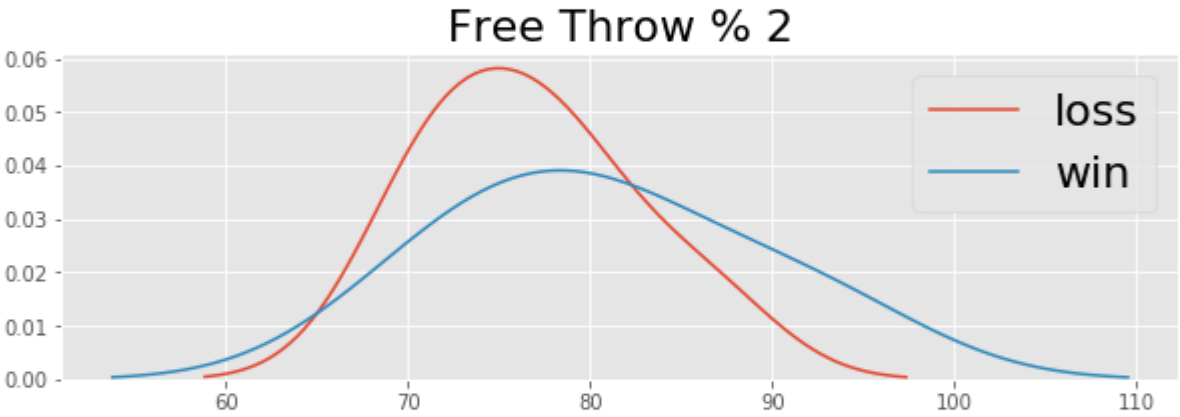

assists 2
loss
win
0.12
0.10
0.08
0.06
0.04
0.02
0.00
15
20
25
30
35
40
45

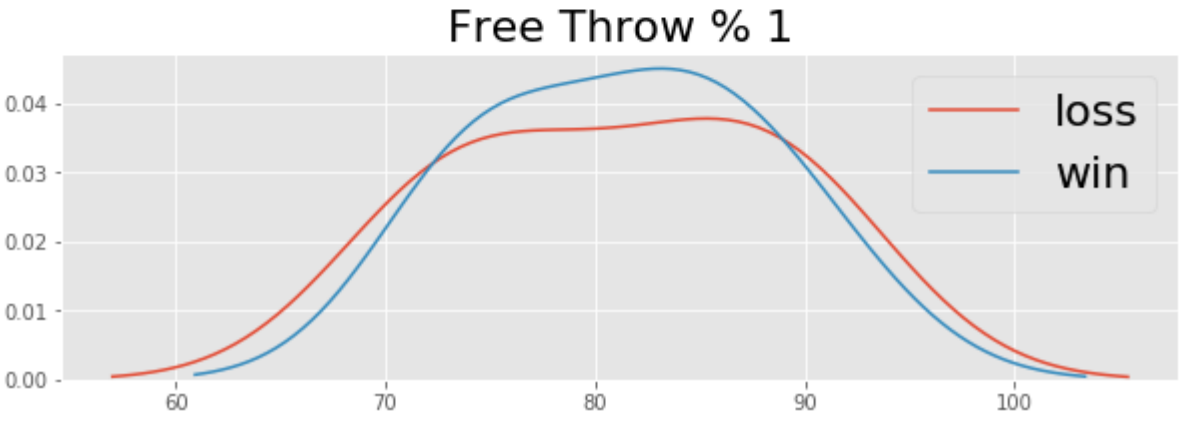

assists 1
loss
win
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
35
40

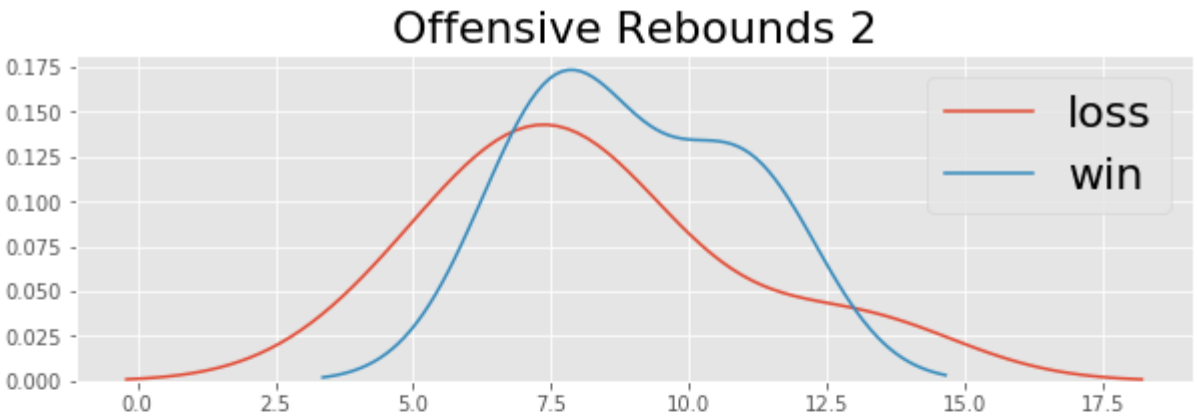

Defensive Rebounds 2
loss
win
0.06
0.05
0.04
0.03
0.02
0.01
0.00
20
30
40
50
60

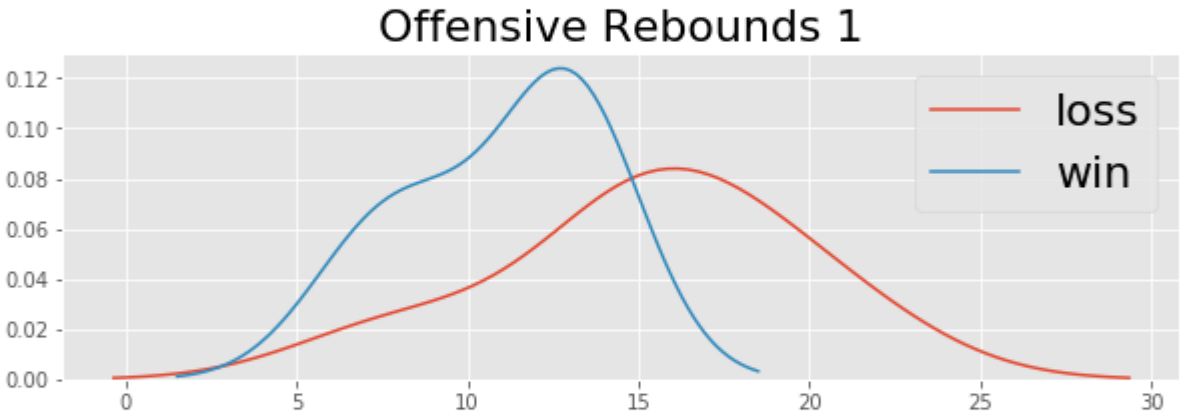

Defensive Rebounds 1
loss
win
0.08
0.06
0.04
0.02
0.00
15
20
25
30
35
40
45
50

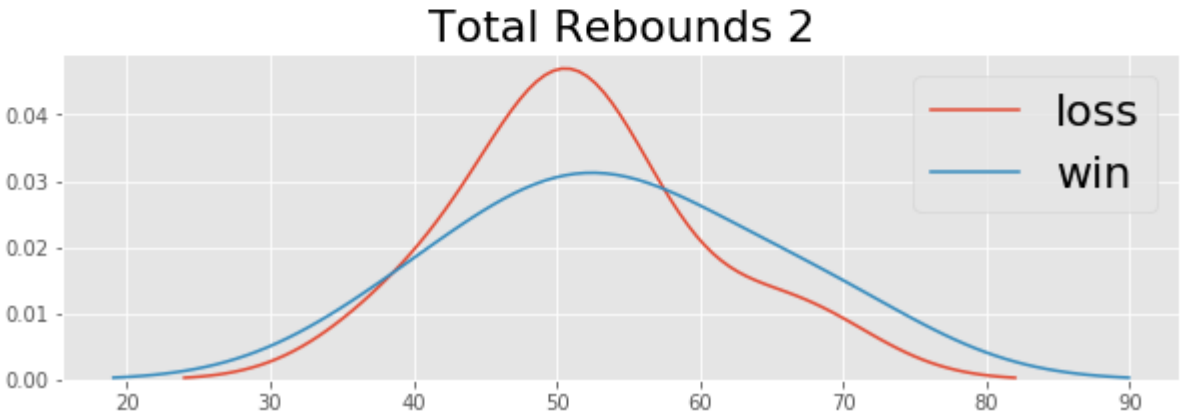

Total Rebounds 1
loss
win
0.06
0.05
0.04
0.03
0.02
0.01
0.00
40
50
60
70
80

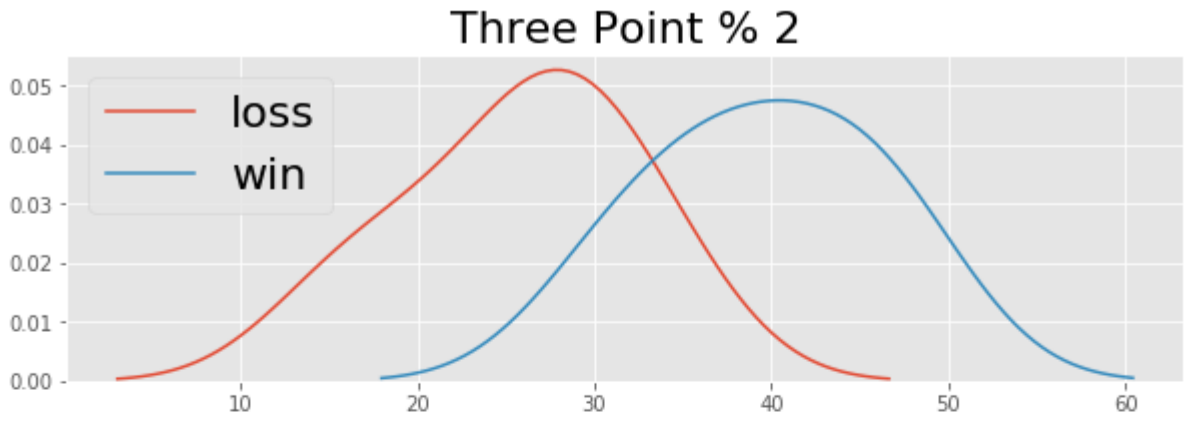

Free Throw % 2
loss
win
0.06
0.05
0.04
0.03
0.02
0.01
0.00
60
70
80
90
100
110

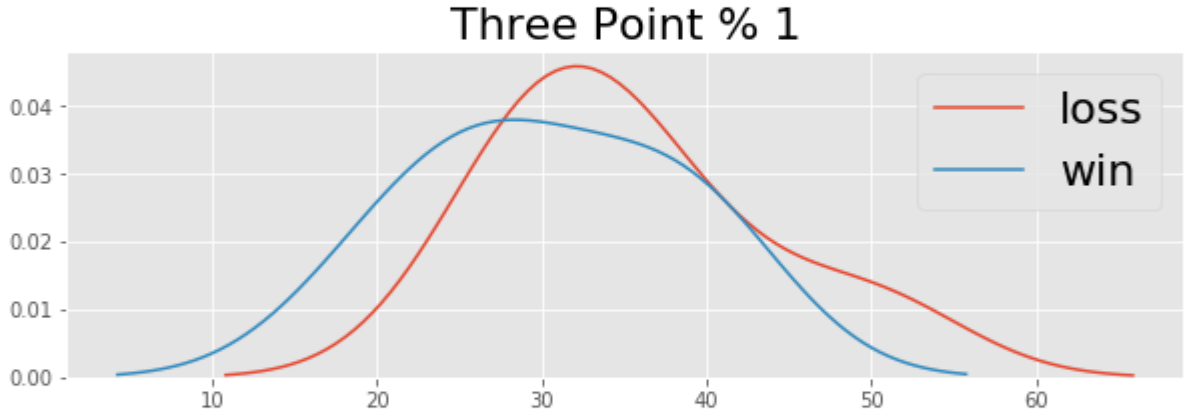

Free Throw % 1
loss
win
0.04
0.03
0.02
0.01
0.00
60
70
80
90
100

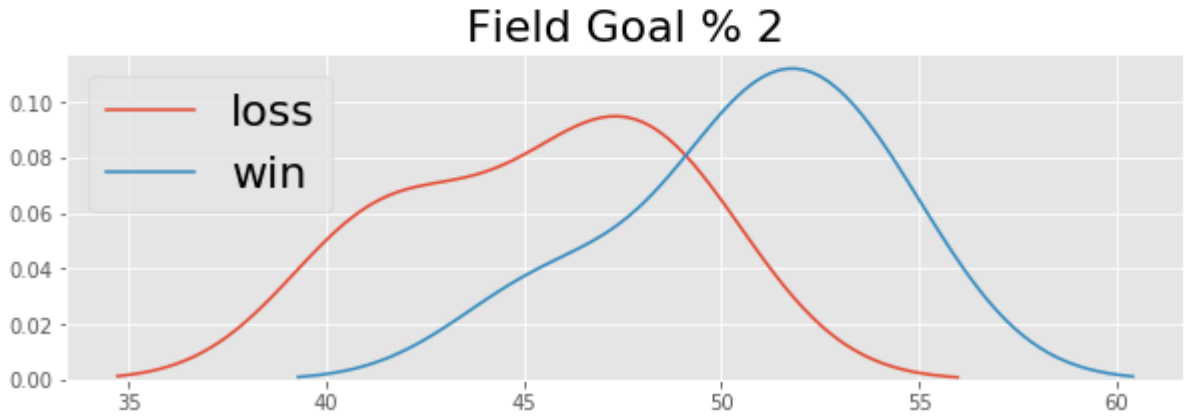

Offensive Rebounds 2
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
loss
win

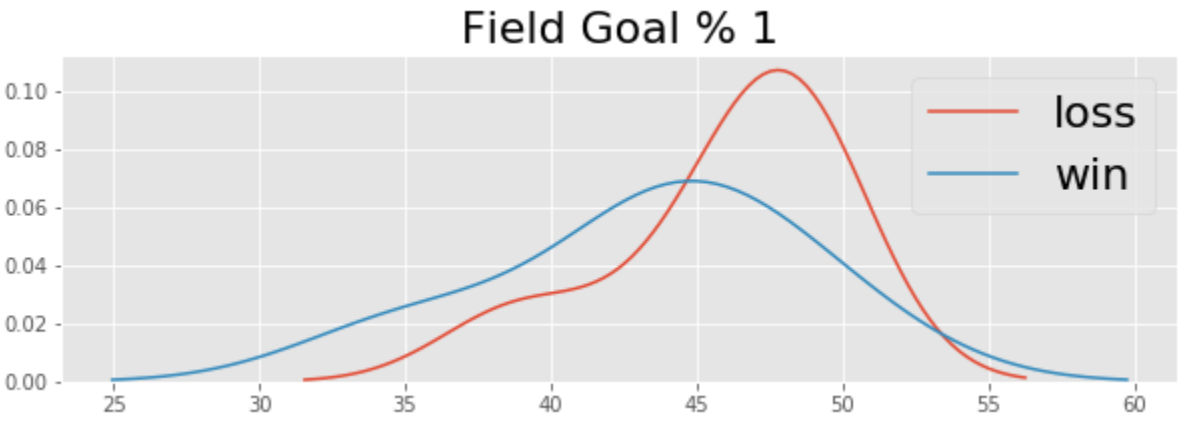

Offensive Rebounds 1
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
loss
win

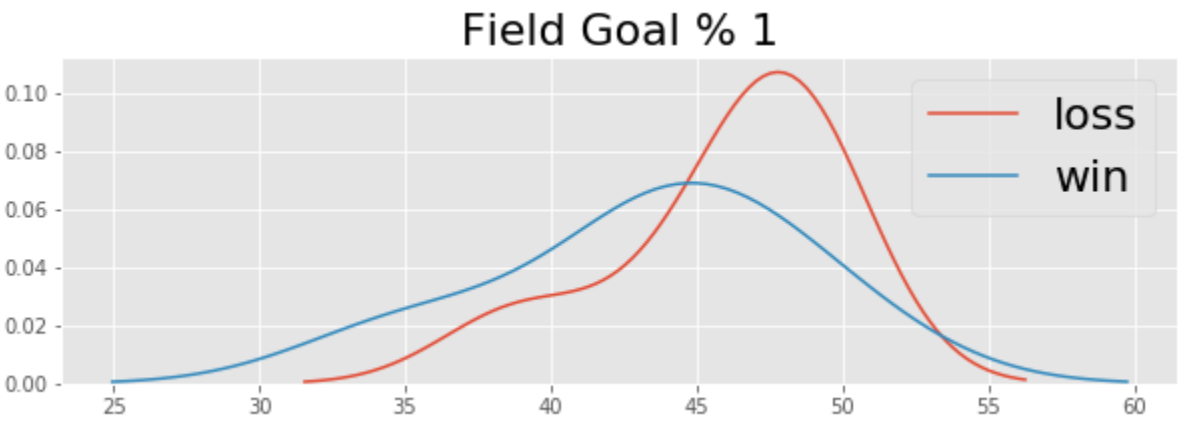

Total Rebounds 2
0.04
0.03
0.02
0.01
0.00
20
30
40
50
60
70
80
90
loss
win


Three Point % 2
loss
win
0.05
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60

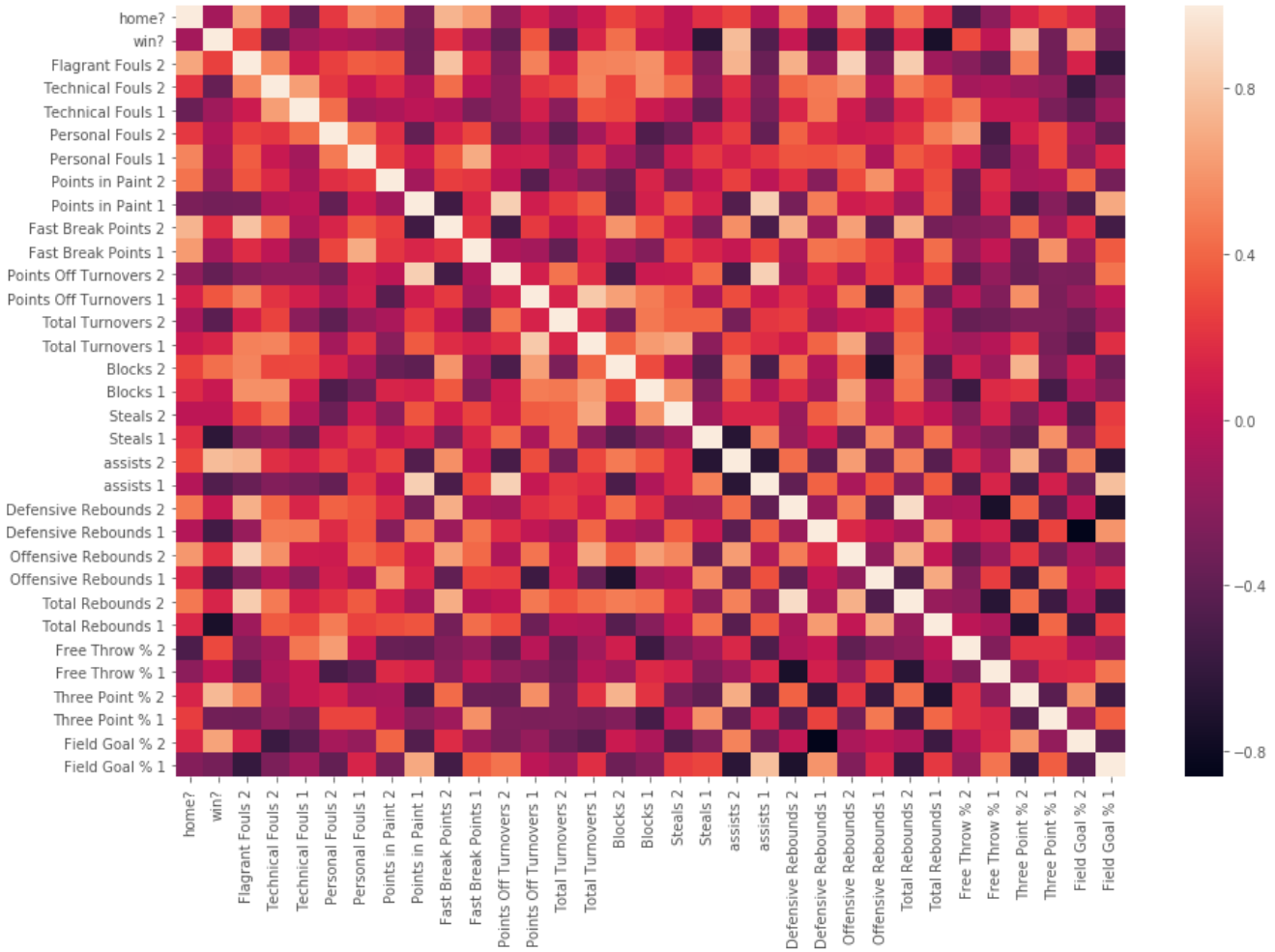

Three Point % 1
loss
win
0.04
0.03
0.02
0.01
0.00
10
20
30
40
50
60

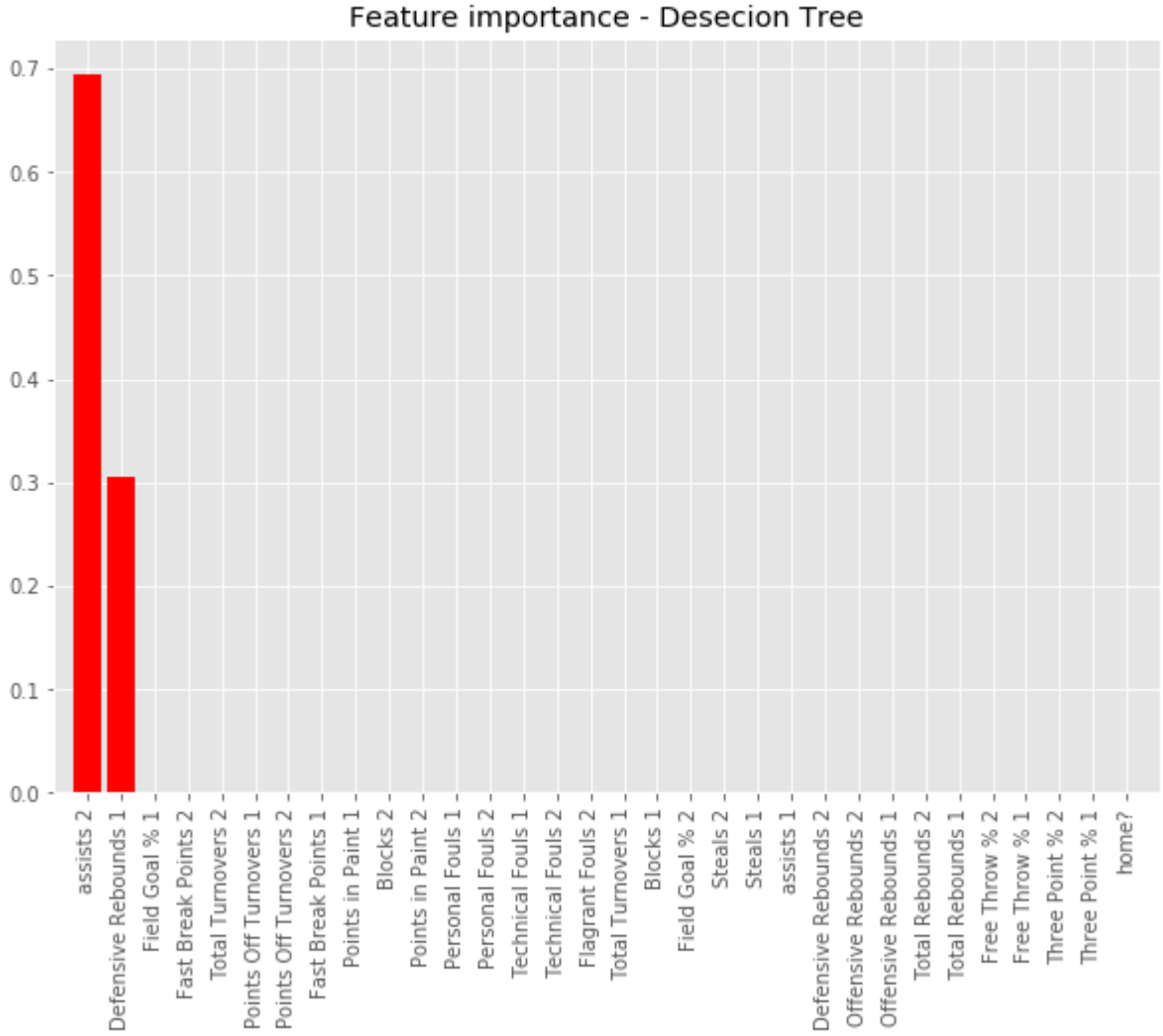

Field Goal % 2
loss
win
0.10
0.08
0.06
0.04
0.02
0.00
35
40
45
50
55
60


Field Goal % 1
0.10
0.08
0.06
0.04
0.02
0.00
25
30
35
40
45
50
55
60
loss
win



Correlation Matrix:




Feature importance - Desecion Tree
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
assists 2
Defensive Rebounds 1
Field Goal % 1
Fast Break Points 2
Total Turnovers 2
Points Off Turnovers 1
Points Off Turnovers 2
Fast Break Points 1
Points in Paint 1
Blocks 2
Points in Paint 2
Personal Fouls 1
Personal Fouls 2
Technical Fouls 1
Technical Fouls 2
Flagrant Fouls 2
Total Turnovers 1
Blocks 1
Field Goal % 2
Steals 2
Steals 1
assists 1
Defensive Rebounds 2
Offensive Rebounds 2
Offensive Rebounds 1
Total Rebounds 2
Total Rebounds 1
Free Throw % 2
Free Throw % 1
Three Point % 2
Three Point % 1
home?