In [ ]:
```
import functions
import os
import warnings
warnings.filterwarnings('ignore')
os.chdir('/Users/stevturn3/Desktop/workspaces/Personal_project/MM')
dir = '/Users/stevturn3/Desktop/workspaces/Personal_project/MM/march-machine
```

Load Data

Here we load relevant data for our analysis

In [ ]:
```
cities = functions.load_df('Cities', dir)
conferences = functions.load_df('Conferences', dir)
tourney_seeds = functions.load_df('2024_tourney_seeds', dir)
tourney_slots = functions.load_df('MNCAATourneySlots', dir)
```

In [ ]:
```
MNCAATourneySeedRoundSlots = functions.load_df('MNCAATourneySlots', dir)
mRegSeasonResults = functions.load_df('MRegularSeasonDetailedResults', dir)
mMassey = functions.load_df('MMasseyOrdinals', dir)
mTourneyResults = functions.load_df('MNCAATourneyDetailedResults', dir)
mTeams = functions.load_df('MTeams', dir)
```

In [ ]:
```
wRegSeasonResults = functions.load_df('WRegularSeasonDetailedResults', dir)
wTourneyResults = functions.load_df('WNCAATourneyDetailedResults', dir)
WNCAATourneySlots = functions.load_df('WNCAATourneySlots', dir)

wTeams = functions.load_df('WTeams', dir)
```

Run Pipeline

We now run the pipeline on both the Men's and Women's Data.

In [ ]:
```
m_r, best_mod_m, m_win_dict = functions.pipeline(mMassey, mRegSeasonResults,
```

In [ ]:
```
wm_r, best_mod_w, w_win_dict = functions.pipeline(None, wRegSeasonResults, W
```

We take a look at the most likely winners over the course of our simulations.

In [ ]:
```
lik = functions.likeli(mTeams, tourney_seeds, 'M', m_r)
print(lik.head(10))
```

```
            TeamName   likeli
Team
W01      Connecticut  0.21195
Z01          Houston  0.09782
Y02        Tennessee  0.09765
X04          Alabama  0.09083
Y01           Purdue  0.08445
Y04           Kansas  0.05346
X07           Dayton  0.04901
X01   North Carolina  0.04538
W02          Iowa St  0.04094
W04           Auburn  0.03963
```

In [ ]:
```
lik = functions.likeli(wTeams, tourney_seeds, 'W', wm_r)
print(lik.head(10))
```

```
            TeamName   likeli
Team
W01   South Carolina  0.24095
Y03              LSU  0.09428
X01            Texas  0.07474
X02          Stanford  0.05838
Z03      Connecticut  0.05731
X04          Gonzaga  0.04109
Y01             Iowa  0.03613
W04          Indiana  0.02924
W02       Notre Dame  0.02603
W03        Oregon St  0.02521
```

Output our results in the form required by the competition.

In [ ]:
```
import pandas as pd
df = pd.concat([m_r, wm_r])
df.to_csv('res.csv')
```

Most Likeli Tourney

Here we take a look at the most likely tournament according to our model. We do so by selecting the team with highest win probability according to our model. It is clear that this will likely favor higher seeded teams, but this will give us an idea of likely upsets, strong teams, etc..

We get an idea of very strong predictors such as Blocks, Steals, Assists and Defensive Rebounds. We also get an idea of the non-normality of the win/loss distribution over many of our features. This indicates we likely shouldn't rely on a normality assumption in our analysis. In fact, it is interesting that many of the distributions are multi-modal. This

is something to keep in mind throughout our analysis.

In [ ]:
```
m_likeli, best_mod_likelim = functions.pipeline_ml(mMassey, mRegSeasonResult
```

In [ ]:
```
r = mTeams.set_index('TeamID').to_dict()['TeamName']
seeds = tourney_seeds[tourney_seeds['Tournament'] == 'M'].set_index('Seed')[
id_from_seeds = {value : key for key,value in seeds.items()}
teamName = [r[seeds[l]] for l in list(m_likeli.Team)]
wins = pd.DataFrame({'Slot' : m_likeli.Slot,'TeamName' : teamName, 'Seed' :
print(wins.to_string())
```

```
    Slot        TeamName Seed
0   R1W1     Connecticut  W01
1   R1W2         Iowa St  W02
2   R1W3        Illinois  W03
3   R1W4          Auburn  W04
4   R1W5    San Diego St  W05
5   R1W6             BYU  W06
6   R1W7           Drake  W10
7   R1W8     FL Atlantic  W08
8   R1X1  North Carolina  X01
9   R1X2         Arizona  X02
10  R1X3          Baylor  X03
11  R1X4         Alabama  X04
12  R1X5    St Mary's CA  X05
13  R1X6      New Mexico  X11
14  R1X7          Dayton  X07
15  R1X8     Michigan St  X09
16  R1Y1          Purdue  Y01
17  R1Y2       Tennessee  Y02
18  R1Y3       Creighton  Y03
19  R1Y4          Kansas  Y04
20  R1Y5         Gonzaga  Y05
21  R1Y6  South Carolina  Y06
22  R1Y7    Colorado St  Y10
23  R1Y8         Utah St  Y08
24  R1Z1         Houston  Z01
25  R1Z2       Marquette  Z02
26  R1Z3        Kentucky  Z03
27  R1Z4            Duke  Z04
28  R1Z5       Wisconsin  Z05
29  R1Z6      Texas Tech  Z06
30  R1Z7        Boise St  Z10
31  R1Z8        Nebraska  Z08
32  R2W1     Connecticut  W01
33  R2W2         Iowa St  W02
34  R2W3             BYU  W06
35  R2W4          Auburn  W04
36  R2X1  North Carolina  X01
```

37	R2X2	Dayton	X07
38	R2X3	Baylor	X03
39	R2X4	Alabama	X04
40	R2Y1	Purdue	Y01
41	R2Y2	Tennessee	Y02
42	R2Y3	Creighton	Y03
43	R2Y4	Kansas	Y04
44	R2Z1	Houston	Z01
45	R2Z2	Marquette	Z02
46	R2Z3	Kentucky	Z03
47	R2Z4	Duke	Z04
48	R3W1	Connecticut	W01
49	R3W2	Iowa St	W02
50	R3X1	Alabama	X04
51	R3X2	Dayton	X07
52	R3Y1	Purdue	Y01
53	R3Y2	Tennessee	Y02
54	R3Z1	Houston	Z01
55	R3Z2	Marquette	Z02
56	R4W1	Connecticut	W01
57	R4X1	Alabama	X04
58	R4Y1	Purdue	Y01
59	R4Z1	Houston	Z01
60	R5WX	Connecticut	W01
61	R5YZ	Purdue	Y01
62	R6CH	Connecticut	W01

The most likely winner is UCONN!