March Madness Basketball Tournament

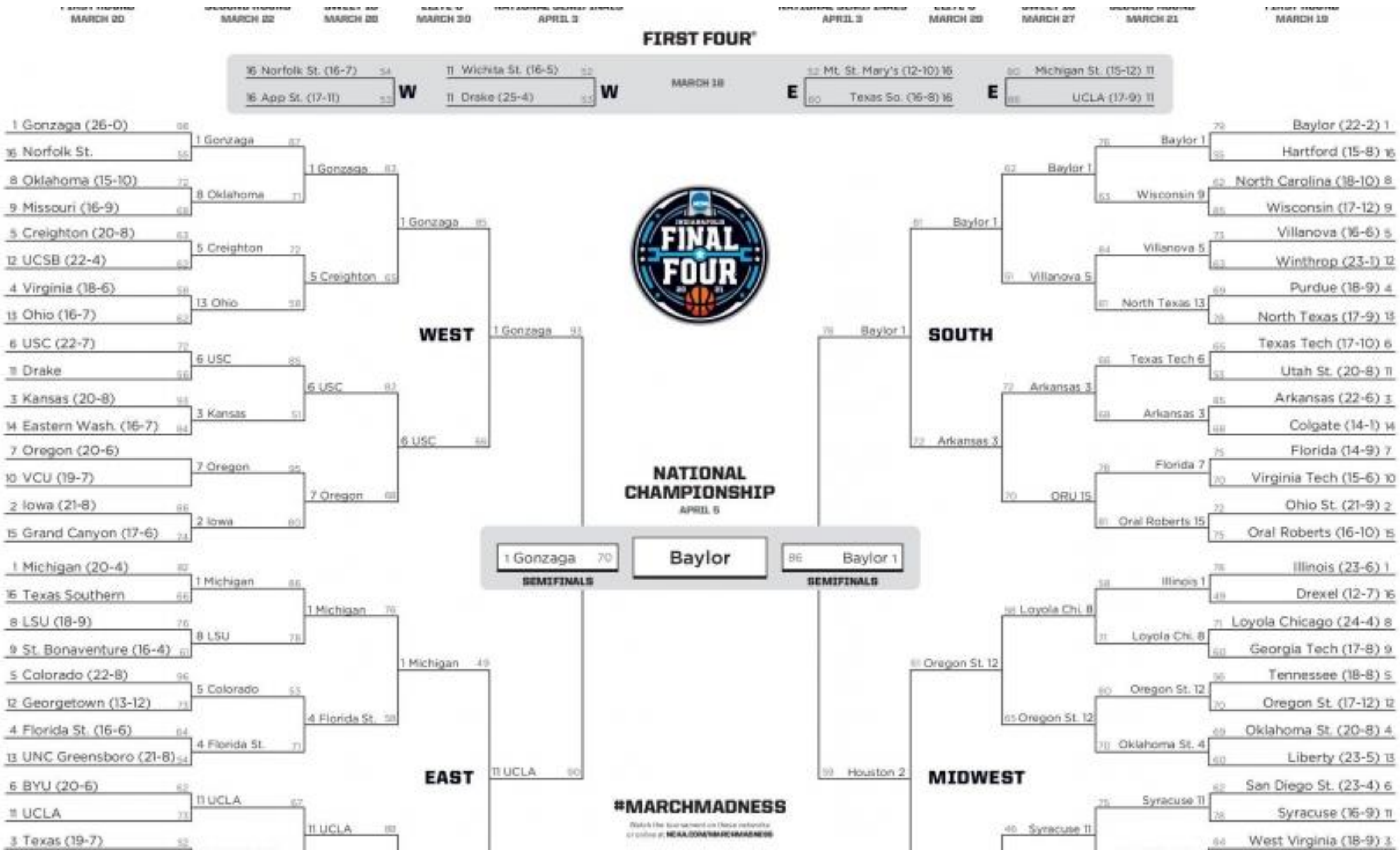
357 Teams in Division 1

20-30 games in the regular season

68 teems seeded for NCAA March Madness

Q : Given regular season data, who is going to win NCAA?

-



FIRST ROUND
MARCH 20

SECOND ROUND
MARCH 22

SWEET 16
MARCH 28

ELITE 8
MARCH 30

NATIONAL SEMIFINALS
APRIL 3

FIRST FOUR®

16 Norfolk St. (16-7) 54
16 App St. (17-11) 53
W

11 Wichita St. (16-5) 52
11 Drake (25-4) 53
W

MARCH 18

52 Mt. St. Mary's (12-10) 16
60 Texas So. (16-8) 16
E

80 Michigan St. (15-12) 11
86 UCLA (17-9) 11
E

NATIONAL SEMIFINALS
APRIL 3

ELITE 8
MARCH 29

SWEET 16
MARCH 27

SECOND ROUND
MARCH 21

FIRST ROUND
MARCH 19

WEST

1 Gonzaga (26-0) 98
16 Norfolk St. 55
8 Oklahoma (15-10) 72
9 Missouri (16-9) 68
5 Creighton (20-8) 63
12 UCSB (22-4) 62
4 Virginia (18-6) 58
13 Ohio (16-7) 62
6 USC (22-7) 72
11 Drake 56
3 Kansas (20-8) 93
14 Eastern Wash. (16-7) 84
7 Oregon (20-6)
10 VCU (19-7)
2 Iowa (21-8) 86
15 Grand Canyon (17-6) 74

1 Gonzaga 87
8 Oklahoma 71
5 Creighton 72
13 Ohio 58
6 USC 85
3 Kansas 51
7 Oregon 95
2 Iowa 80

1 Gonzaga 83
5 Creighton 65
6 USC 82
7 Oregon 68

1 Gonzaga 85
6 USC 66

1 Gonzaga 93

EAST

1 Michigan (20-4) 82
16 Texas Southern 66
8 LSU (18-9) 76
9 St. Bonaventure (16-4) 61
5 Colorado (22-8) 96
12 Georgetown (13-12) 73
4 Florida St. (16-6) 64
13 UNC Greensboro (21-8) 54
6 BYU (20-6) 62
11 UCLA 73
3 Texas (19-7) 52

1 Michigan 86
8 LSU 78
5 Colorado 53
4 Florida St. 71
11 UCLA 67

1 Michigan 76
4 Florida St. 58
11 UCLA 88

1 Michigan 49

11 UCLA 90

NATIONAL CHAMPIONSHIP
APRIL 5

1 Gonzaga 70
SEMIFINALS

Baylor

86 Baylor 1
SEMIFINALS

#MARCHMADNESS

Watch the tournament on these networks or online at NCAA.COM/MARCHMADNESS

SOUTH

79 Baylor (22-2) 1
55 Hartford (15-8) 16
62 North Carolina (18-10) 8
85 Wisconsin (17-12) 9
73 Villanova (16-6) 5
63 Winthrop (23-1) 12
69 Purdue (18-9) 4
78 North Texas (17-9) 13
65 Texas Tech (17-10) 6
53 Utah St. (20-8) 11
85 Arkansas (22-6) 3
68 Colgate (14-1) 14
75 Florida (14-9) 7
70 Virginia Tech (15-6) 10
72 Ohio St. (21-9) 2
75 Oral Roberts (16-10) 15

76 Baylor 1
63 Wisconsin 9
84 Villanova 5
61 North Texas 13
66 Texas Tech 6
68 Arkansas 3
78 Florida 7
81 Oral Roberts 15

62 Baylor 1
51 Villanova 5
72 Arkansas 3
70 ORU 15

81 Baylor 1
72 Arkansas 3

78 Baylor 1

MIDWEST

78 Illinois (23-6) 1
49 Drexel (12-7) 16
71 Loyola Chicago (24-4) 8
60 Georgia Tech (17-8) 9
56 Tennessee (18-8) 5
70 Oregon St. (17-12) 12
69 Oklahoma St. (20-8) 4
60 Liberty (23-5) 13
62 San Diego St. (23-4) 6
78 Syracuse (16-9) 11
84 West Virginia (18-9) 3

58 Illinois 1
71 Loyola Chi. 8
80 Oregon St. 12
70 Oklahoma St. 4
75 Syracuse 11

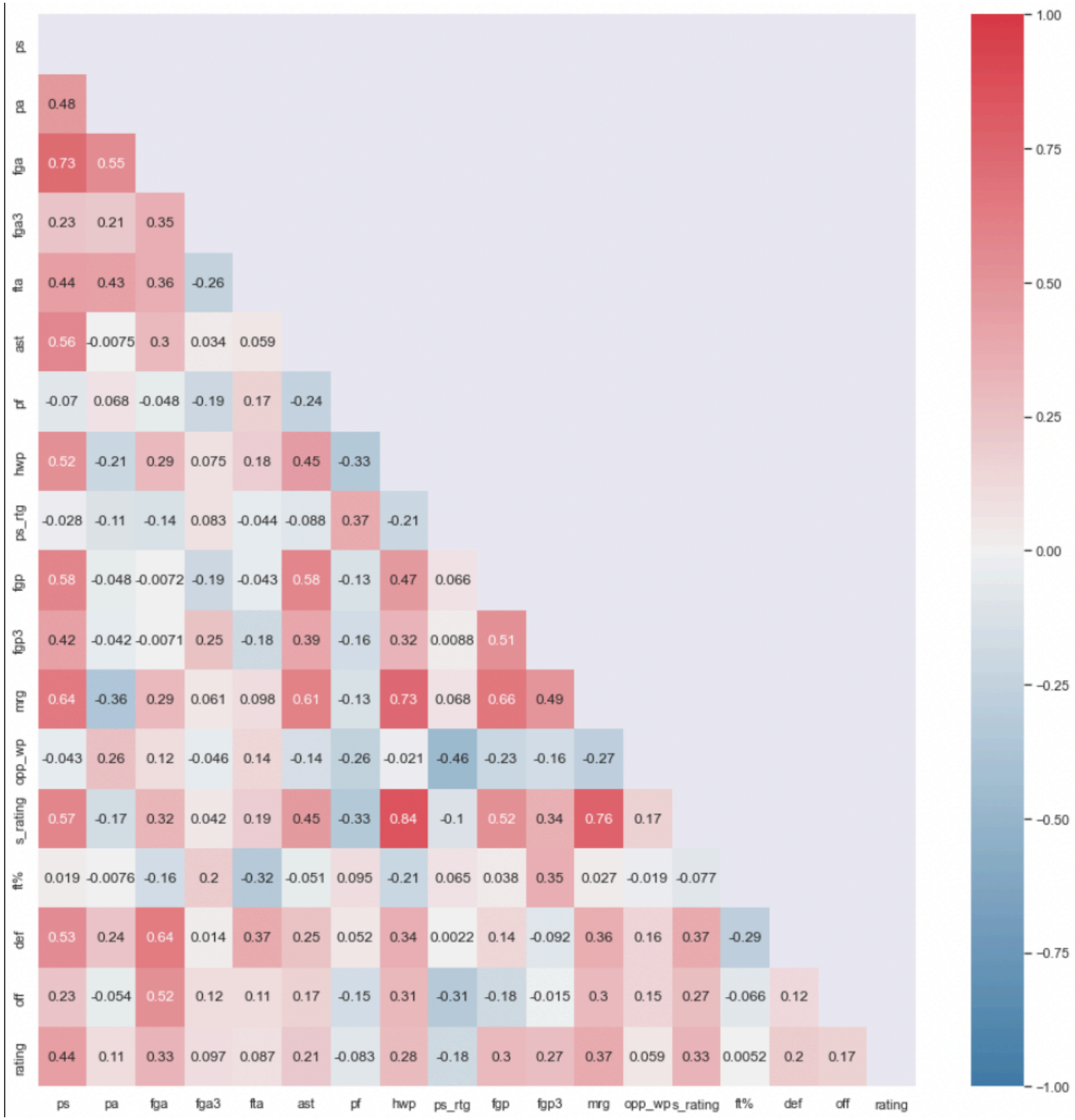
58 Loyola Chi. 8
65 Oregon St. 12
46 Syracuse 11

61 Oregon St. 12

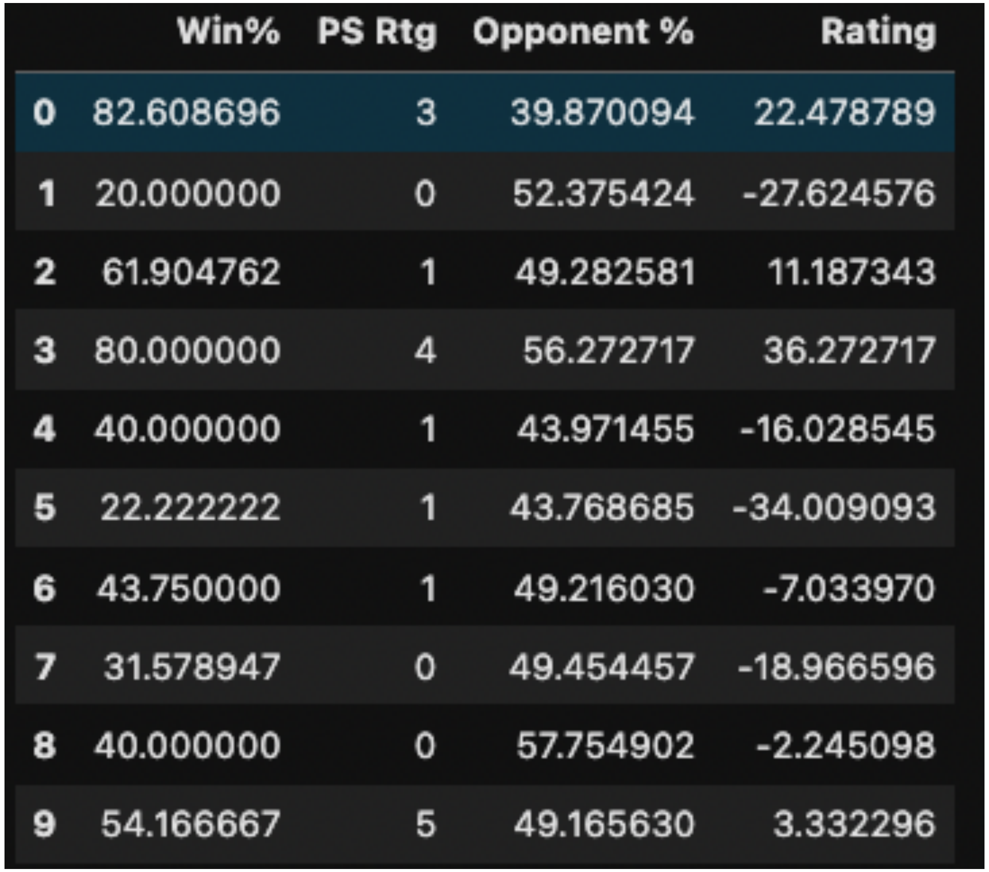
59 Houston 2

Challanges to start with

- We had structured, but vast data (20 CSV files)
- Each team plays with a different set of teams.
- Post season tournements are indicative but small in sample size
- How to account for recency and fixture difficulty

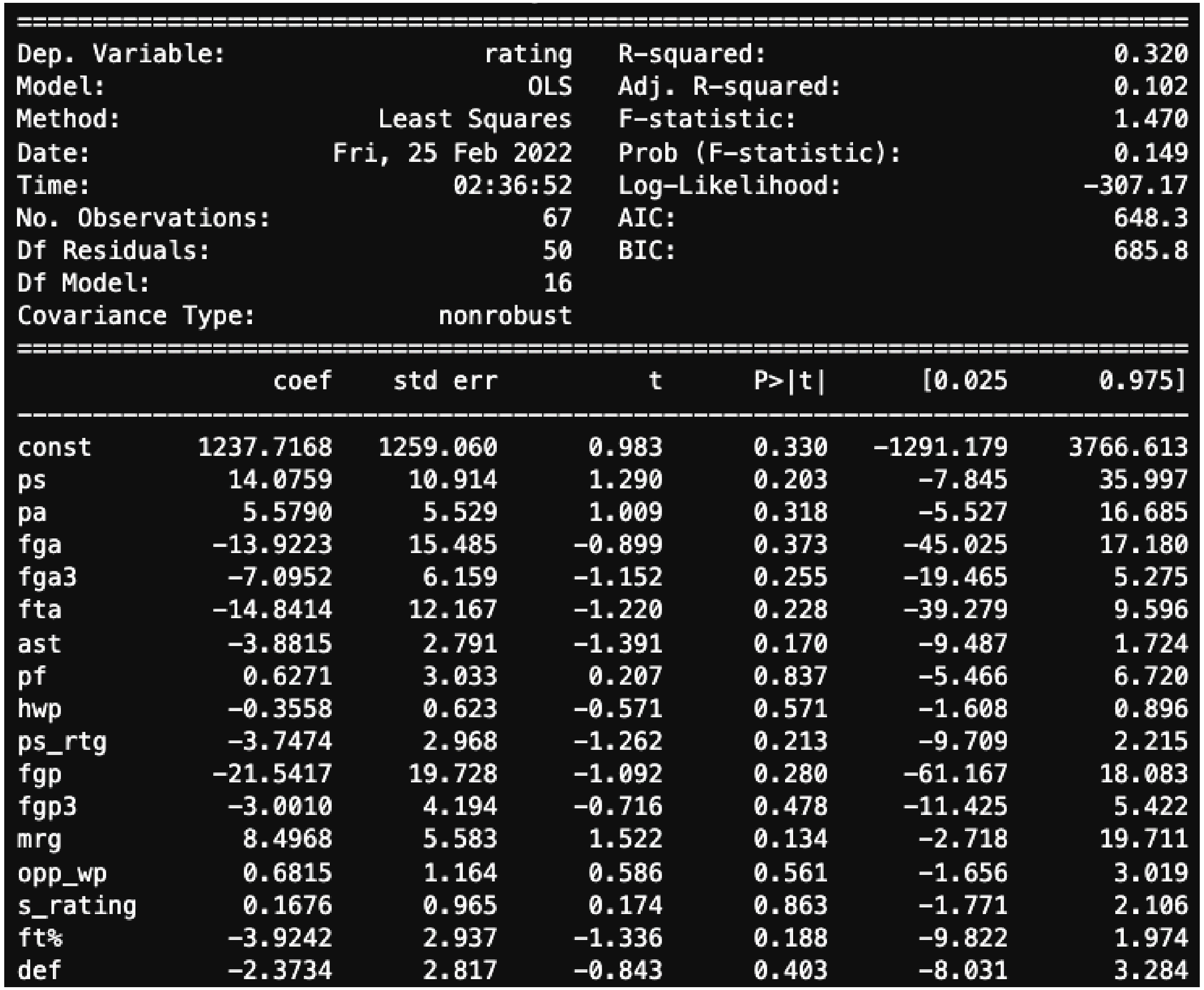

ps
pa 0.48
fga 0.73 0.55
fga3 0.23 0.21 0.35
fta 0.44 0.43 0.36 -0.26
ast 0.56 -0.0075 0.3 0.034 0.059
pf -0.07 0.068 -0.048 -0.19 0.17 -0.24
hwp 0.52 -0.21 0.29 0.075 0.18 0.45 -0.33
ps_rtg -0.028 -0.11 -0.14 0.083 -0.044 -0.088 0.37 -0.21
fgp 0.58 -0.048 -0.0072 -0.19 -0.043 0.58 -0.13 0.47 0.066
fgp3 0.42 -0.042 -0.0071 0.25 -0.18 0.39 -0.16 0.32 0.0088 0.51
mrg 0.64 -0.36 0.29 0.061 0.098 0.61 -0.13 0.73 0.068 0.66 0.49
opp_wp -0.043 0.26 0.12 -0.046 0.14 -0.14 -0.26 -0.021 -0.46 -0.23 -0.16 -0.27
s_rating 0.57 -0.17 0.32 0.042 0.19 0.45 -0.33 0.84 -0.1 0.52 0.34 0.76 0.17
ft% 0.019 -0.0076 -0.16 0.2 -0.32 -0.051 0.095 -0.21 0.065 0.038 0.35 0.027 -0.019 -0.077
def 0.53 0.24 0.64 0.014 0.37 0.25 0.052 0.34 0.0022 0.14 -0.092 0.36 0.16 0.37 -0.29
off 0.23 -0.054 0.52 0.12 0.11 0.17 -0.15 0.31 -0.31 -0.18 -0.015 0.3 0.15 0.27 -0.066 0.12
rating 0.44 0.11 0.33 0.097 0.087 0.21 -0.083 0.28 -0.18 0.3 0.27 0.37 0.059 0.33 0.0052 0.2 0.17
ps pa fga fga3 fta ast pf hwp ps_rtg fgp fgp3 mrg opp_wp s_rating ft% def off rating
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Challanges to start with



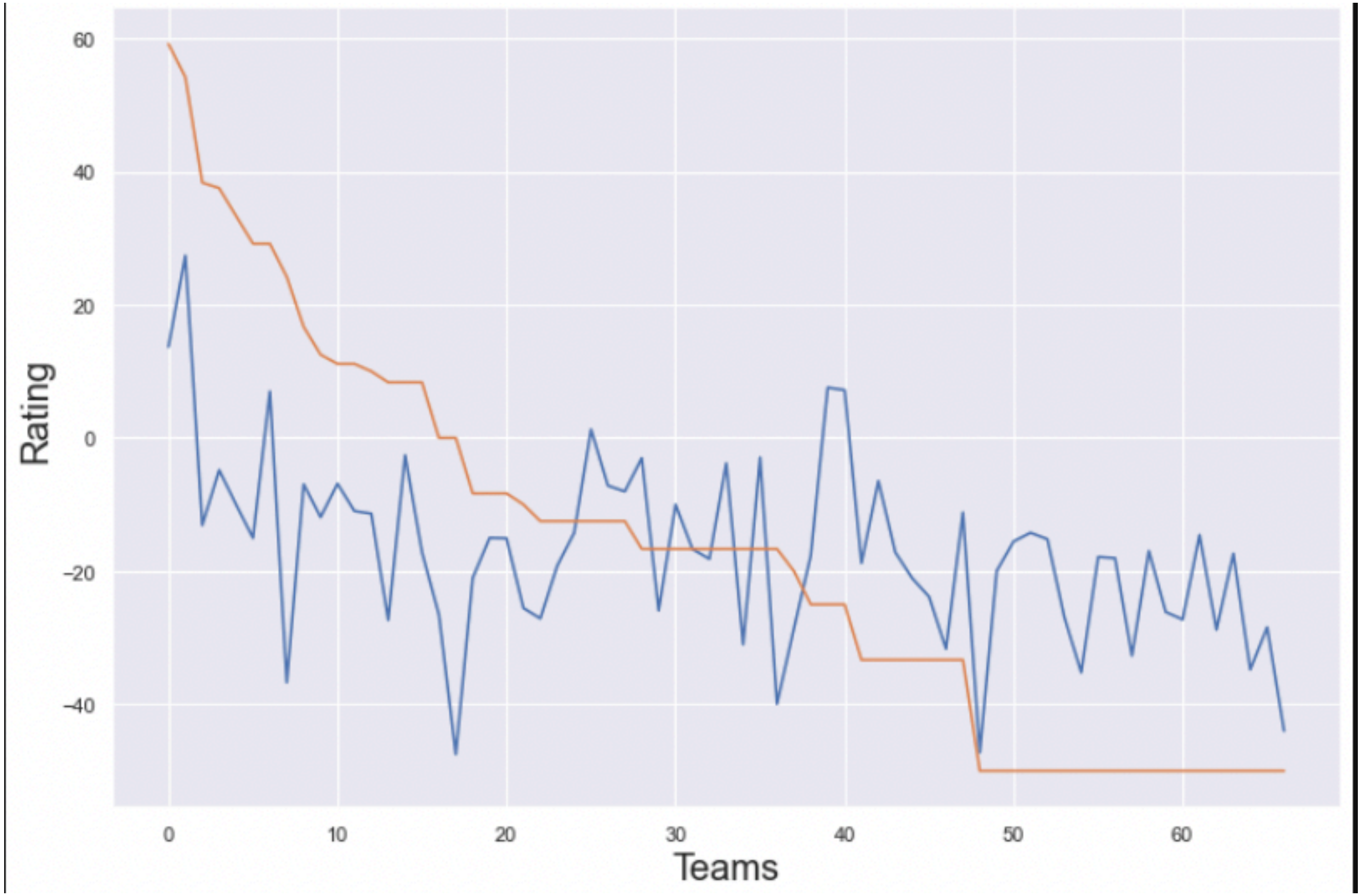
	Win%	PS Rtg	Opponent %	Rating
0	82.608696	3	39.870094	22.478789
1	20.000000	0	52.375424	-27.624576
2	61.904762	1	49.282581	11.187343
3	80.000000	4	56.272717	36.272717
4	40.000000	1	43.971455	-16.028545
5	22.222222	1	43.768685	-34.009093
6	43.750000	1	49.216030	-7.033970
7	31.578947	0	49.454457	-18.966596
8	40.000000	0	57.754902	-2.245098
9	54.166667	5	49.165630	3.332296

Regression with almost all variables

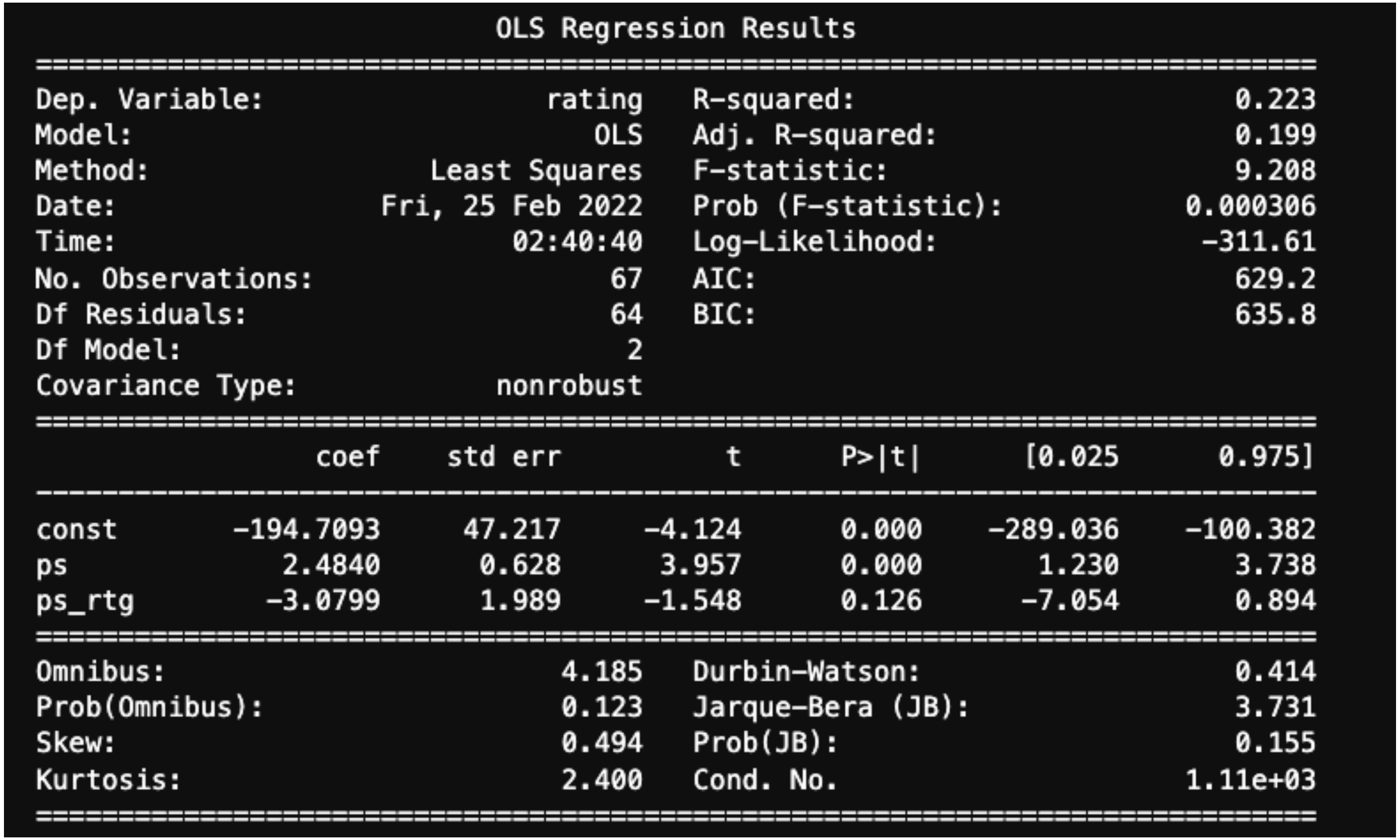


```
==============================================================================
Dep. Variable:                 rating   R-squared:                       0.320
Model:                            OLS   Adj. R-squared:                  0.102
Method:                 Least Squares   F-statistic:                     1.470
Date:                Fri, 25 Feb 2022   Prob (F-statistic):              0.149
Time:                        02:36:52   Log-Likelihood:                -307.17
No. Observations:                  67   AIC:                             648.3
Df Residuals:                      50   BIC:                             685.8
Df Model:                          16
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1237.7168   1259.060      0.983      0.330   -1291.179    3766.613
ps            14.0759     10.914      1.290      0.203      -7.845      35.997
pa             5.5790      5.529      1.009      0.318      -5.527      16.685
fga          -13.9223     15.485     -0.899      0.373     -45.025      17.180
fga3          -7.0952      6.159     -1.152      0.255     -19.465       5.275
fta          -14.8414     12.167     -1.220      0.228     -39.279       9.596
ast           -3.8815      2.791     -1.391      0.170      -9.487       1.724
pf             0.6271      3.033      0.207      0.837      -5.466       6.720
hwp           -0.3558      0.623     -0.571      0.571      -1.608       0.896
ps_rtg        -3.7474      2.968     -1.262      0.213      -9.709       2.215
fgp          -21.5417     19.728     -1.092      0.280     -61.167      18.083
fgp3          -3.0010      4.194     -0.716      0.478     -11.425       5.422
mrg            8.4968      5.583      1.522      0.134      -2.718      19.711
opp_wp         0.6815      1.164      0.586      0.561      -1.656       3.019
s_rating       0.1676      0.965      0.174      0.863      -1.771       2.106
ft%           -3.9242      2.937     -1.336      0.188      -9.822       1.974
def           -2.3734      2.817     -0.843      0.403      -8.031       3.284
```

Fit of the Model

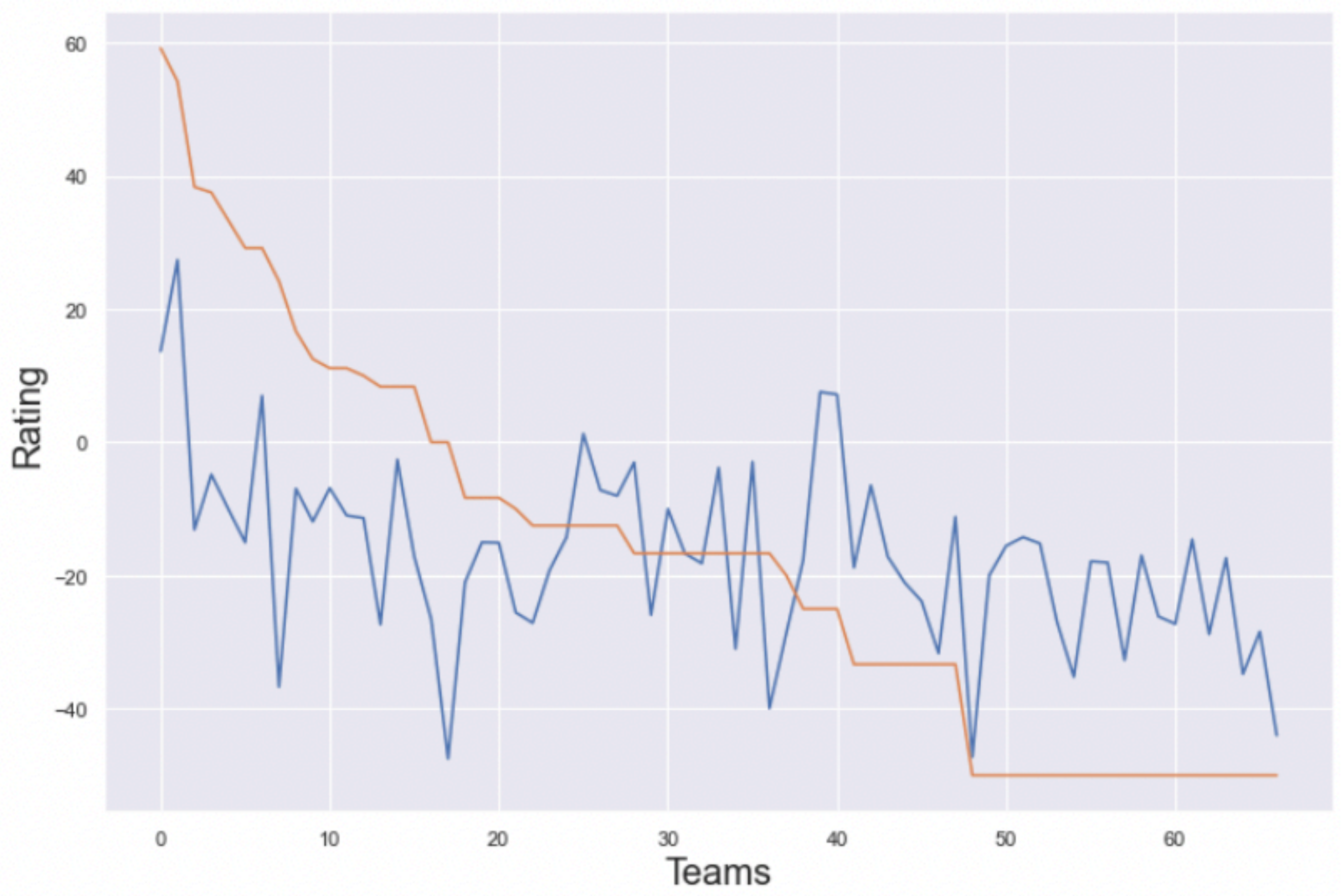


Lowest AIC Regression



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 rating   R-squared:                       0.223
Model:                            OLS   Adj. R-squared:                  0.199
Method:                 Least Squares   F-statistic:                     9.208
Date:                Fri, 25 Feb 2022   Prob (F-statistic):           0.000306
Time:                        02:40:40   Log-Likelihood:                -311.61
No. Observations:                  67   AIC:                             629.2
Df Residuals:                      64   BIC:                             635.8
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -194.7093     47.217     -4.124      0.000    -289.036    -100.382
ps             2.4840      0.628      3.957      0.000       1.230       3.738
ps_rtg        -3.0799      1.989     -1.548      0.126      -7.054       0.894
==============================================================================
Omnibus:                        4.185   Durbin-Watson:                   0.414
Prob(Omnibus):                  0.123   Jarque-Bera (JB):                3.731
Skew:                           0.494   Prob(JB):                        0.155
Kurtosis:                       2.400   Cond. No.                     1.11e+03
==============================================================================
```

Fit of the Model (Lowest AIC)



How did it go / what did we find:

Turns out that College Basketball is not so easily predictable:

- NCAA Tournament results are hard to interpret
- Teams play only against a small portion of all teams, so having (un)favorable match-ups skew the perceived strength during the „Season“

Conclusions:

- The fact that we didn‘t find any obvious correlation doesn‘t mean there can‘t be some – actually there is groups that have been using models based on Machine Learning tools in the past, which predicted the results quite well (be it thanks to luck, or model quality)
- Would be cool to look further into this, but currently outside our skill range