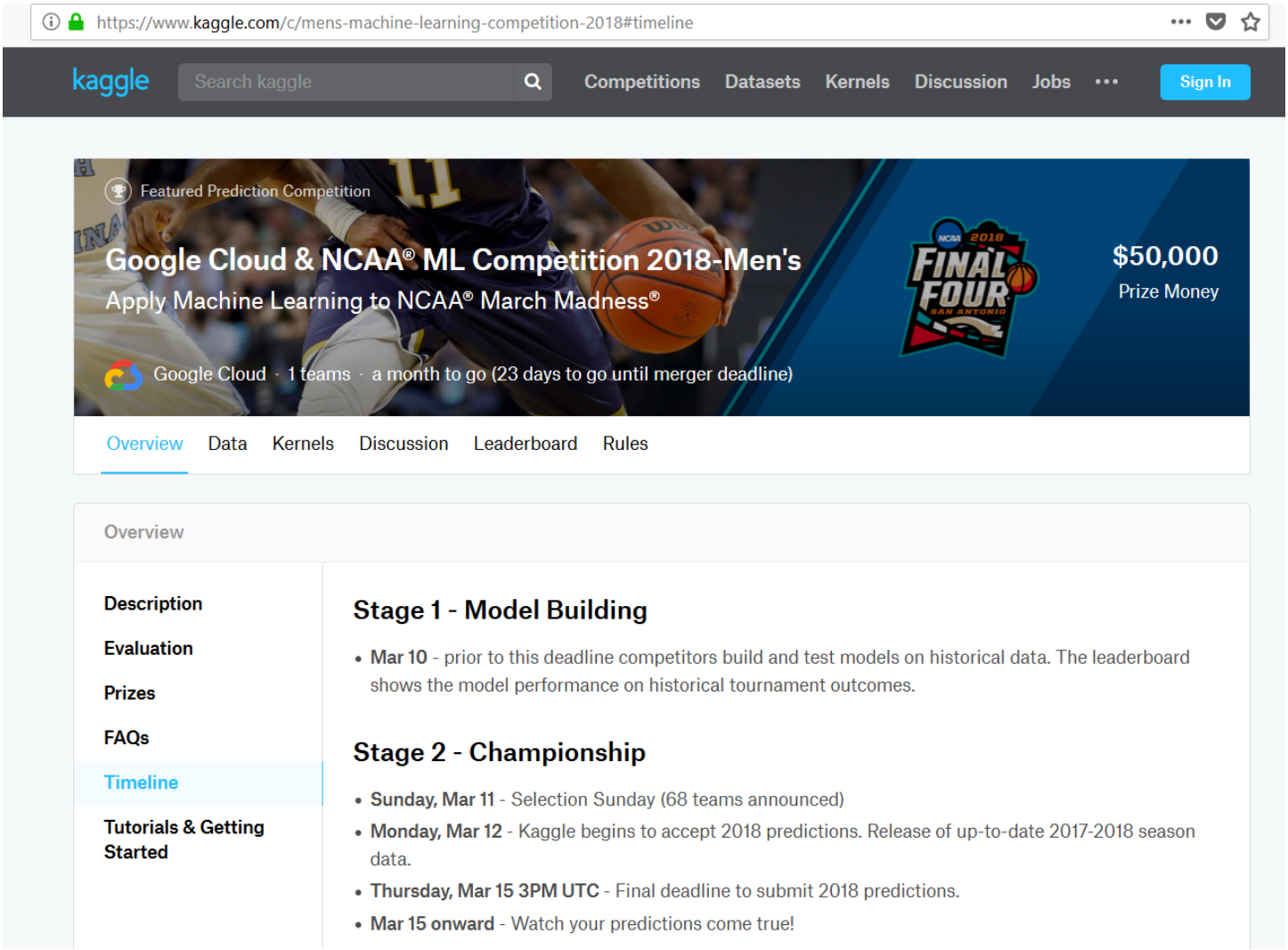
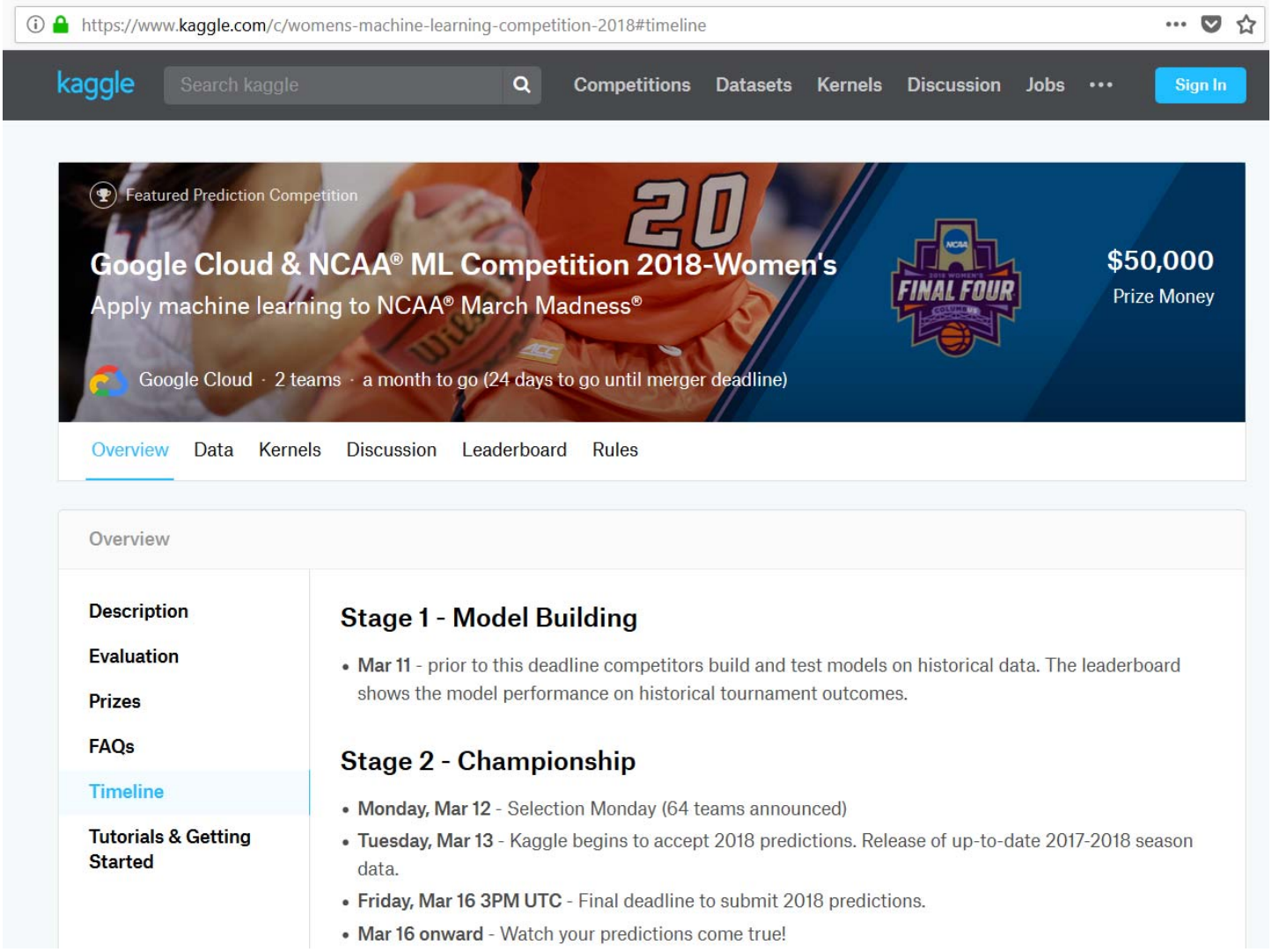
I would like to provide an overview of the way our contest works, first because there are two distinct stages to the contest, and second because the way that you submit your predictions and the way that we score the accuracy of your predictions, is different between our contest and a traditional March Madness bracket competition.

Let's start by reviewing the overall timelines:



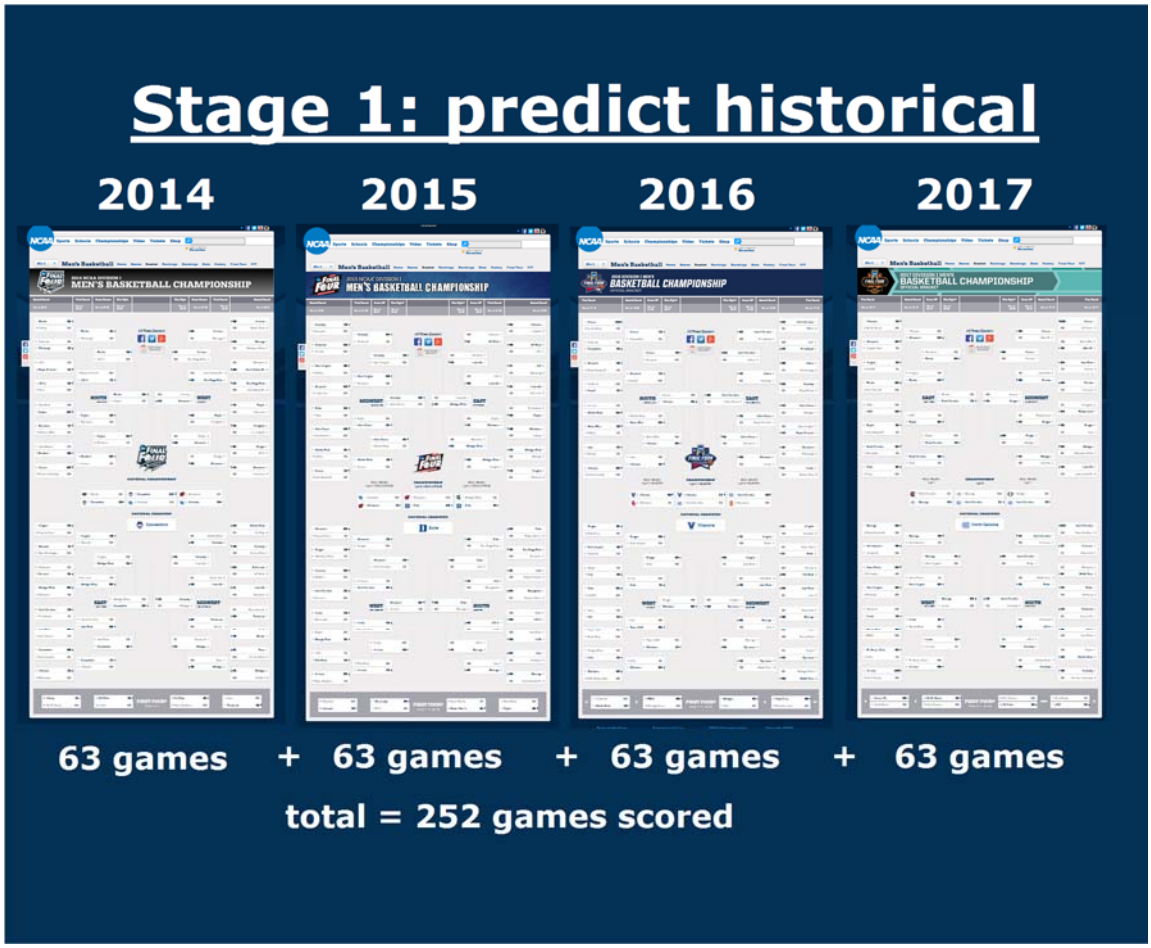
The contest is divided into two stages. Stage 1 for the men's tournament involves predictions for past years, and completes on March 10, right before Selection Sunday. Stage 2 involves predictions for the current year, and begins shortly after the tournament field and the bracket pairings are announced. The final Stage 2 deadline for submission is shortly before the round of 64 begins on the morning of Thursday, March 15.

And the women's contest timeline is shifted later by one day:



Stage 1 for the women's tournament completes on March 11, right before the women's field is announced on March 12, otherwise known as Selection Monday. Then Stage 2 begins shortly after the pairings are announced, and the final Stage 2 deadline for submission of predictions for the women's tournament is shortly before the round of 64 begins on the morning of Friday, March 16.

During Stage 1, you will make predictions for the last four tournaments - the tournaments that were played in March 2014, March 2015, March 2016, and March 2017:



Each submission will be a full list of predicted winning percentages for all possible games across those four tournaments, and you will be scored based on how accurate your predictions were, when scored against the final 63 games each year, for a total of 252 games.

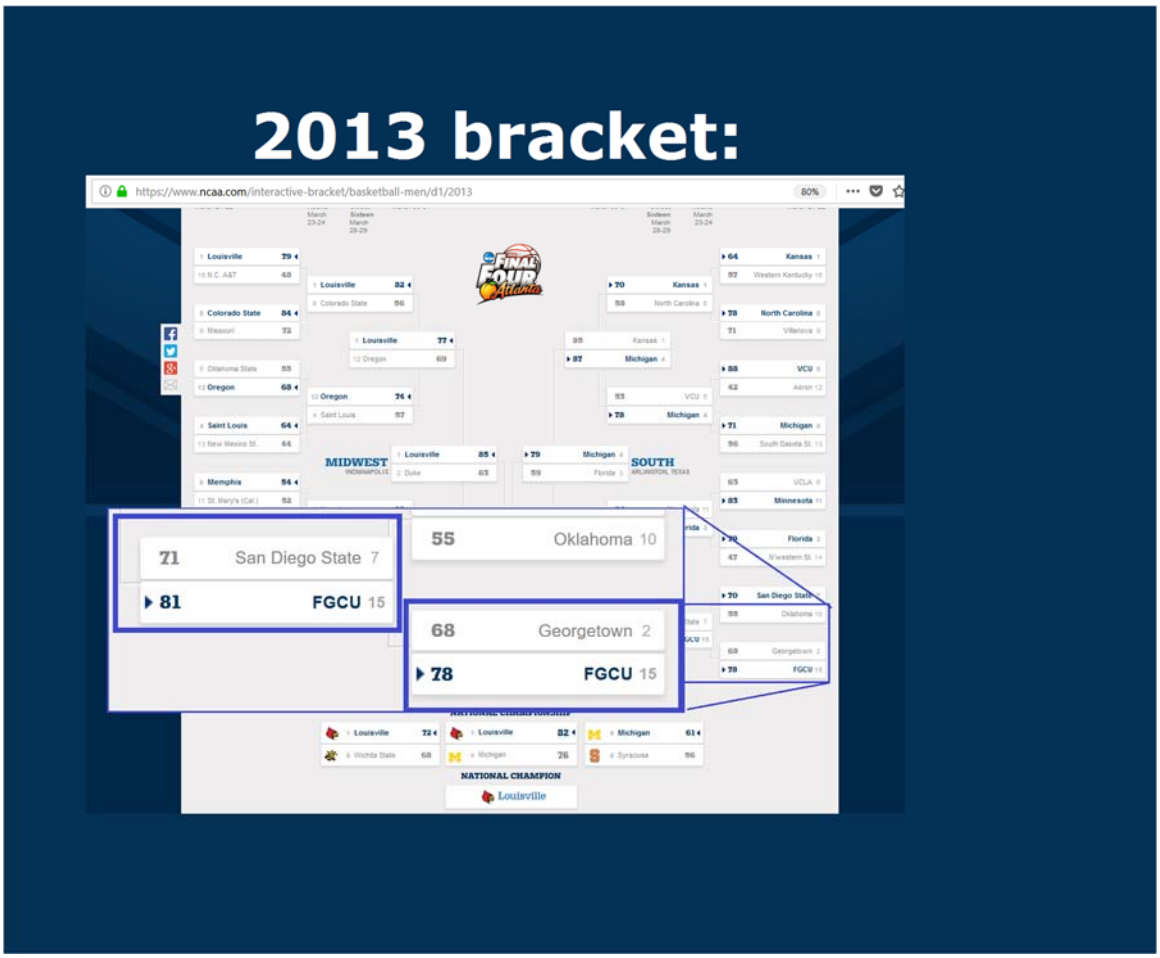
Now of course, there is an immediate and obvious question here. In the same way that I could go online and grab these screen shots of the completed tournament bracket, anyone could go look up the publicly-available final scores of these games that were played in the past. And in fact the scores themselves are readily available in the contest data files as well. Without much trouble, it is possible to prepare a perfect Stage 1 submission file with zero error in the predictions, that will score a perfect 0.00 on the leaderboard. People do this every year, and we have to remove their submissions from the Stage 1 leaderboard. So what's the point of Stage 1?

Because this data is publicly available, there is no prize money and no Kaggle rankings/points awarded for doing well at Stage 1. Nevertheless, it is a very important part of the contest. Stage 1 is where you can develop your predictive model, and double-check your understanding of the submission format and the scoring function, by making submissions to the Stage 1 Leaderboard and confirming your leaderboard score. It also encourages you in the right direction for how to build a good model. You want a prediction approach that not only makes sense in the present, but would have worked well in the past at predicting games from previous tournaments. So it is very useful to know how well your approach would have worked in recent years.

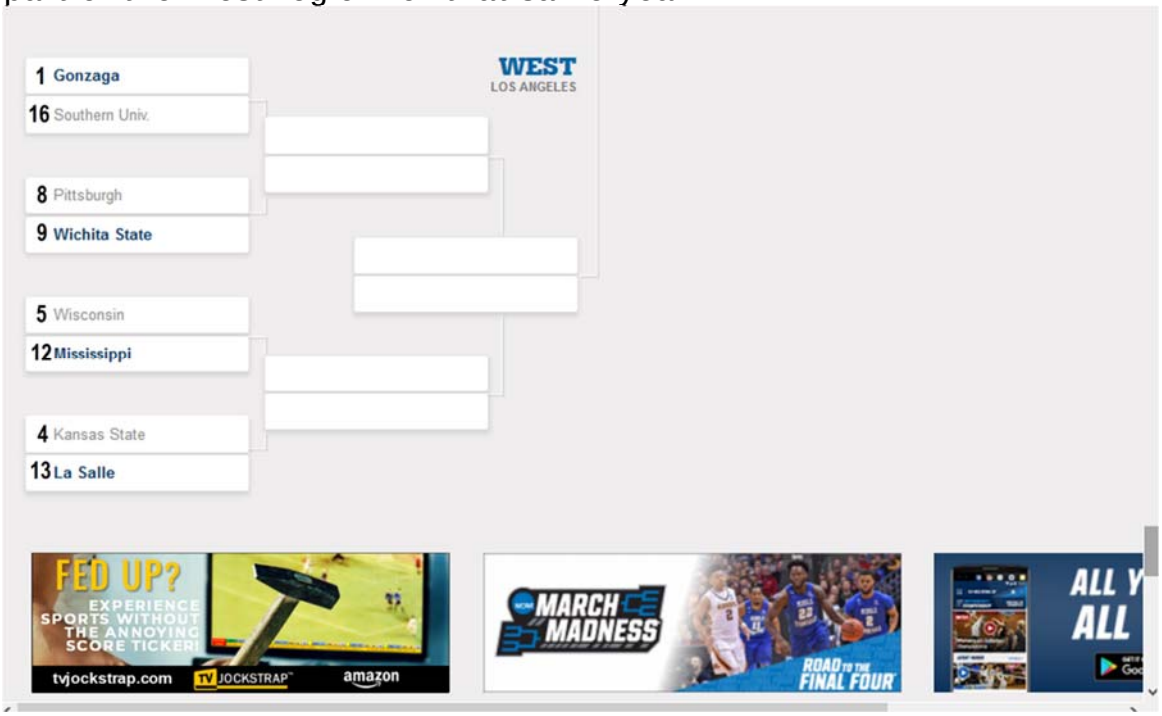
Because the real prize comes if you can predict well in Stage 2.

For stage 2, you will submit predicted winning percentages (before the tournament) for all possible matchups in the current 2018 tournament, whether it be the women's tournament or the men's tournament. The accuracy of those 2018 tournament predictions will be scored by looking at the 63 games that were actually played in the tournament, and seeing whose submission did the best job at predicting those winners, and the prize money will be awarded accordingly to the top three finishers.

In addition to those multiple contest stages, it is important to understand how our scoring differs from a traditional March Madness bracket competition. To illustrate this, let's drill into the details of a particular year's bracket, in this case the 2013 men's tournament.

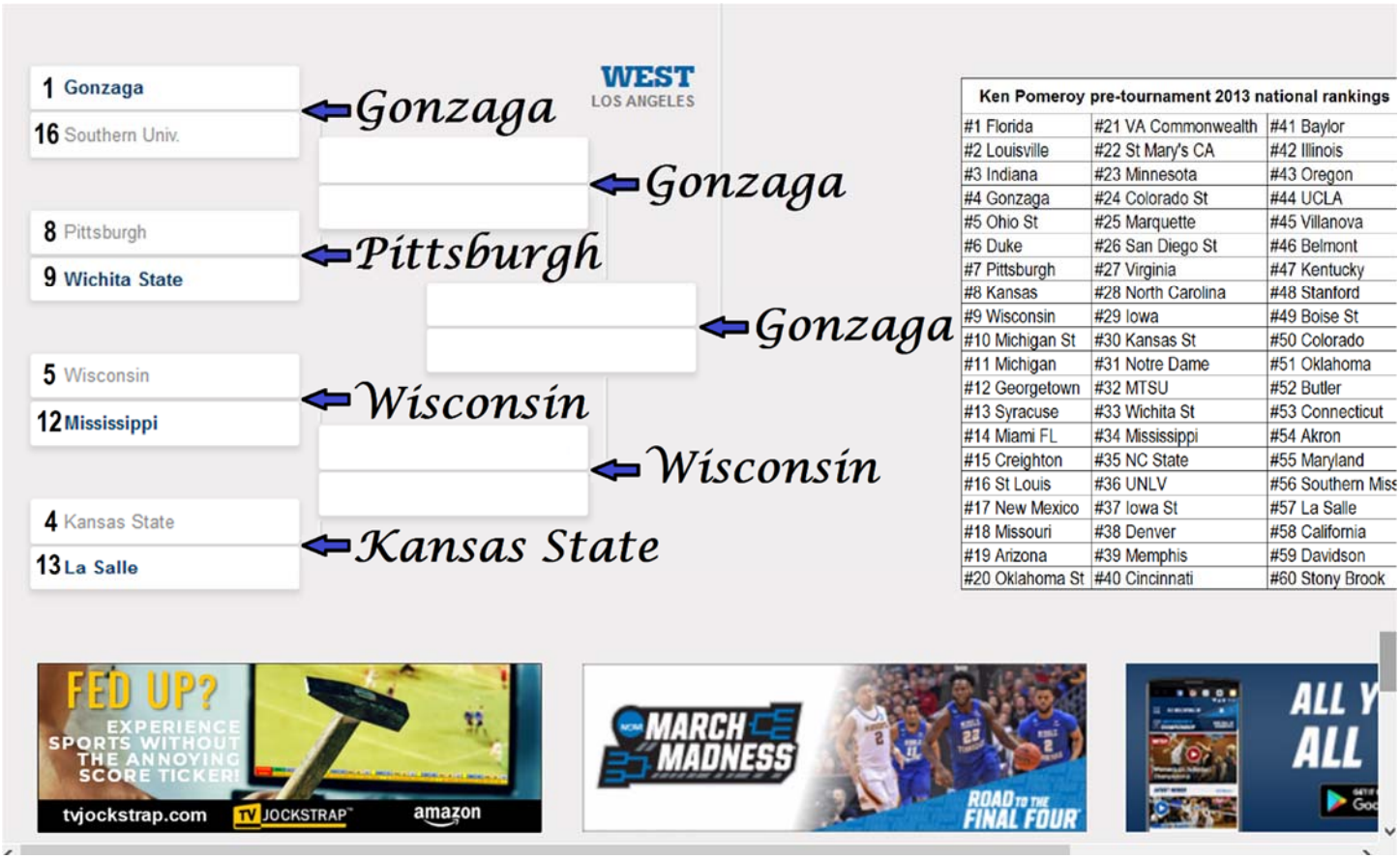


This tournament was famous for having a whole bunch of upsets in the early rounds, most notably with Florida Gulf Coast, who had only been a Division-I program for five years, nevertheless winning its first two games as the #15 seed in the South Region, becoming the only #15 seed to ever reach the Sweet Sixteen. However, let's look at part of the West region for that same year:



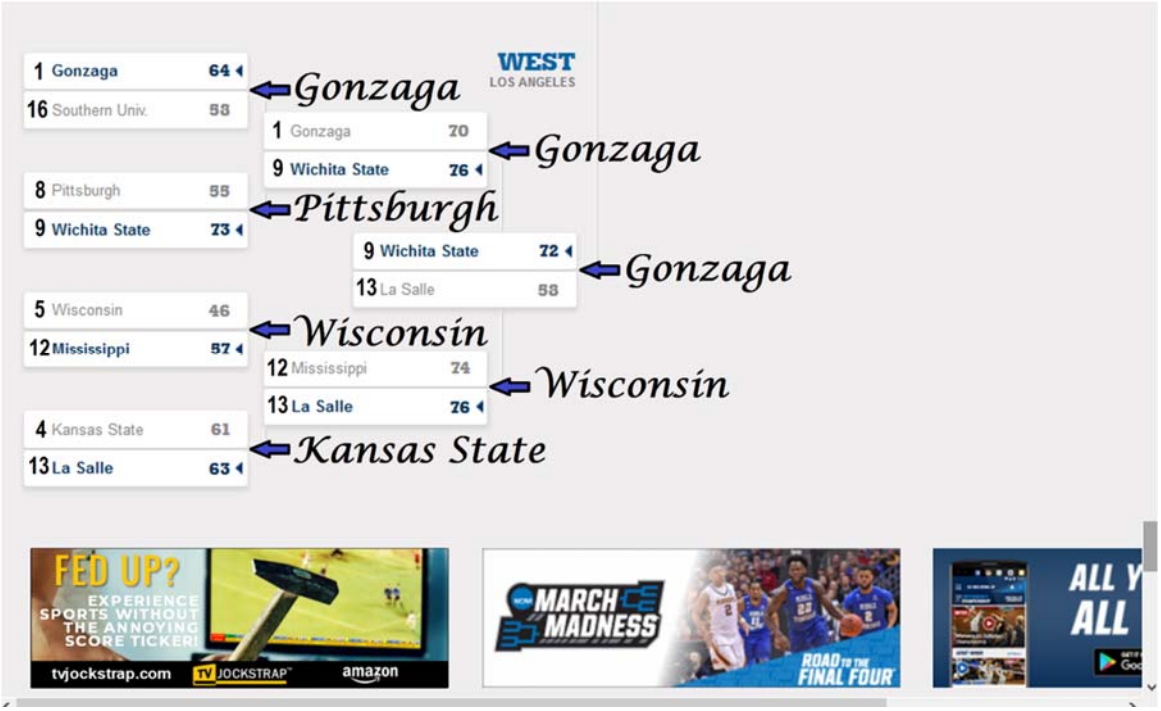
In a traditional bracket competition, you are presented with an empty bracket and you need to fill in your predicted winning team in each bracket slot. Here is one half of the West Region in that 2013 men's tournament. To make our predictions for a traditional bracket competition, let's say we are relying on Ken Pomeroy's pre-tournament national rankings from 2013 to guide us.

Guided by those Pomeroy rankings, we fill in our bracket. For the most part, we are predicting the strongest seeds will advance, so we have predictions that the #1, #8, #4, and #5 seeds will win in the first round, and in the second round we predict a slight upset of 5-seed Wisconsin beating 4-seed Kansas State, but otherwise it's a straightforward “chalk” prediction of the strongest seeds advancing.

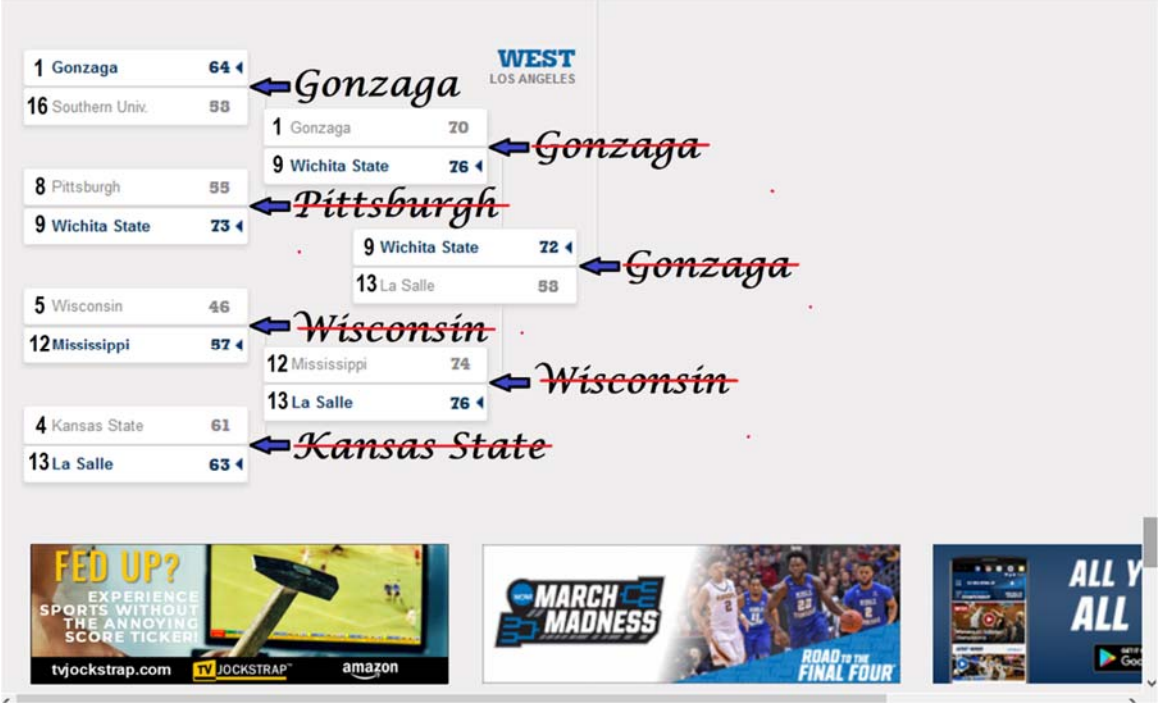


Ken Pomeroy pre-tournament 2013 national rankings		
#1 Florida	#21 VA Commonwealth	#41 Baylor
#2 Louisville	#22 St Mary's CA	#42 Illinois
#3 Indiana	#23 Minnesota	#43 Oregon
#4 Gonzaga	#24 Colorado St	#44 UCLA
#5 Ohio St	#25 Marquette	#45 Villanova
#6 Duke	#26 San Diego St	#46 Belmont
#7 Pittsburgh	#27 Virginia	#47 Kentucky
#8 Kansas	#28 North Carolina	#48 Stanford
#9 Wisconsin	#29 Iowa	#49 Boise St
#10 Michigan St	#30 Kansas St	#50 Colorado
#11 Michigan	#31 Notre Dame	#51 Oklahoma
#12 Georgetown	#32 MTSU	#52 Butler
#13 Syracuse	#33 Wichita St	#53 Connecticut
#14 Miami FL	#34 Mississippi	#54 Akron
#15 Creighton	#35 NC State	#55 Maryland
#16 St Louis	#36 UNLV	#56 Southern Miss
#17 New Mexico	#37 Iowa St	#57 La Salle
#18 Missouri	#38 Denver	#58 California
#19 Arizona	#39 Memphis	#59 Davidson
#20 Oklahoma St	#40 Cincinnati	#60 Stony Brook

Then when the games are actually played, a traditional bracket competition would count up all the times you correctly predicted the team that won each slot.

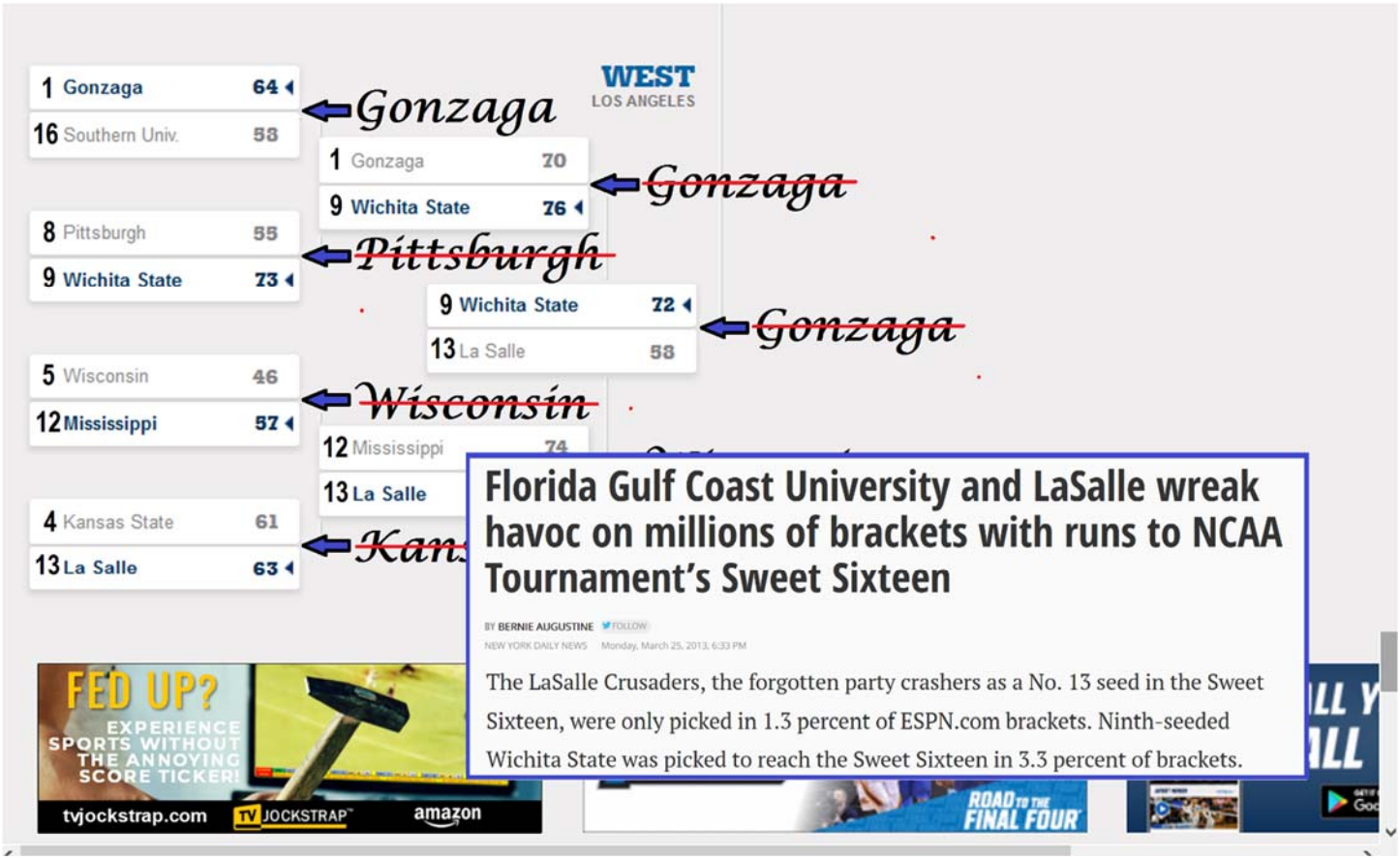


So we would cross out the predictions for each slot that didn't identify the team that won, and a traditional bracket competition would score those on a per-slot basis:



This part of the 2013 bracket is clearly "busted" at historical levels, since the teams predicted to advance far didn't actually get very far at all, with an extremely unlikely Sweet Sixteen matchup between Wichita State and La Salle. In fact, this is the only time there has ever been a game between a #9 seed and a #13 seed.

Now of course, this kind of thing happens to a lesser degree all the time, and that's part of the fun, to see where the upsets happen and who predicted them. But it brings up an interesting point about what you are being asked to predict. If you look at this Sweet Sixteen game, between #9 and #13, this is a game that virtually nobody foresaw in their bracket, and so virtually nobody (pre-tournament) really was thinking directly about what might happen if La Salle and Wichita State faced each other in Round 3.



Analysis of ESPN's bracket submissions indicates that only one bracket in thirty would have had Wichita State winning their first two games and getting into the Sweet Sixteen, and only one bracket in eighty would have had La Salle getting this far. So probably less than one person in two thousand had this particular matchup predicted to happen in the Sweet Sixteen, and thus almost nobody had directly predicted who would win a game between Wichita State and La Salle (the game was actually won by Wichita State).

That's in a traditional bracket competition.

In the case of the Kaggle contest, every single contest submission would have directly predicted what would happen in a potential Sweet Sixteen game between Wichita State and La Salle, no matter how unlikely this matchup seemed. Because in the Kaggle contest, you are asked to predict what would happen in every possible matchup between two tournament teams, just in case it does happen and we need to score the accuracy of those predictions.

To continue the example, we can use the simple Sonas-developed formulas to convert from Pomeroy's national rankings into predicted winning percentage:

PowerRating = 100.0 - 2.32*LN(Rank + 1) - Rank/25.3 - (Rank/205.0)^2
RatingDiff = difference in power ratings of the two teams
WinPct = 1.0 / (1.0 + {POWER(10.0, RatingDiff/12.0)))

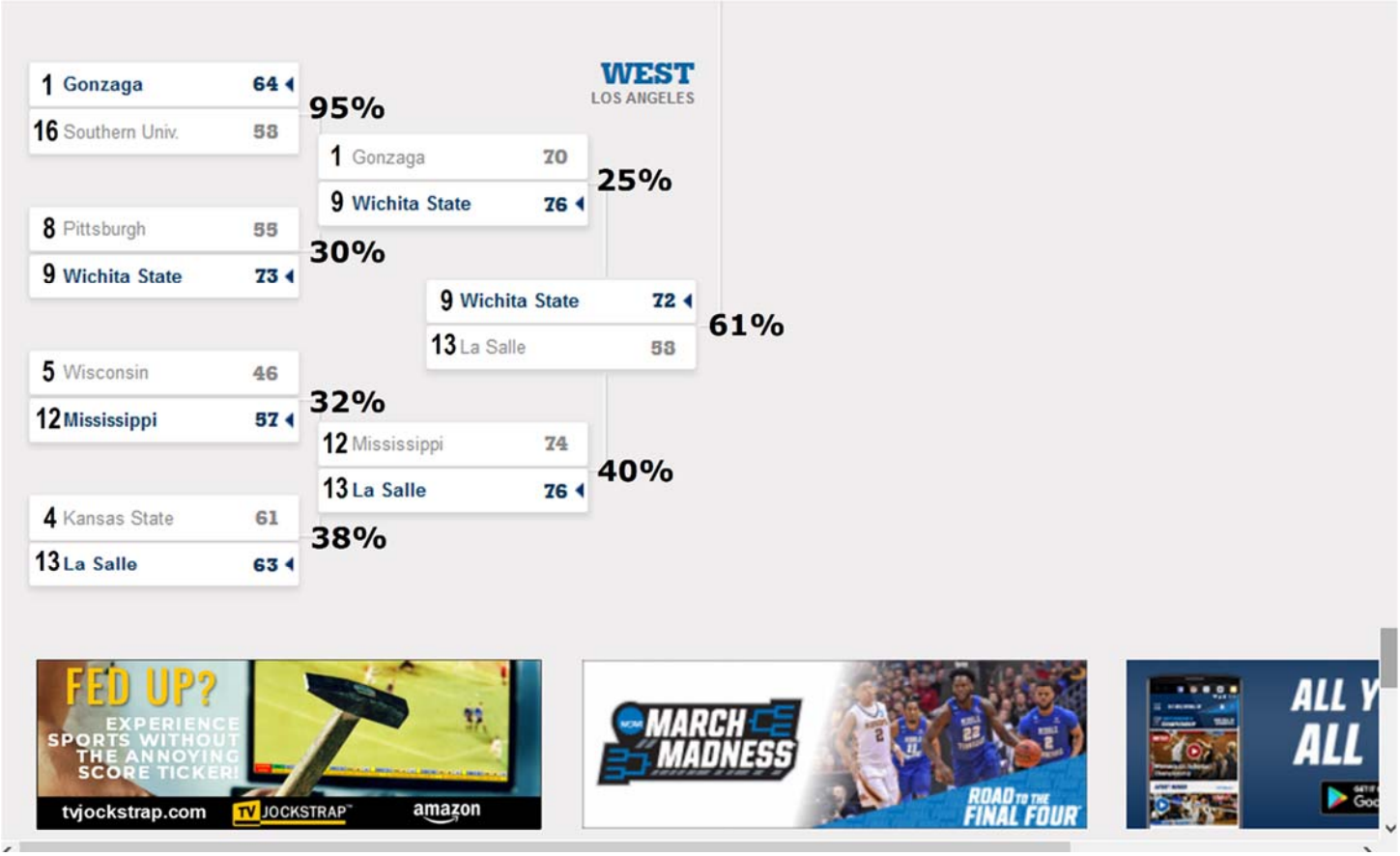
And these calculations allow us to make some winning percentage predictions, using those 2013 Pomeroy national rankings published before the tournament started, for all possible matchups in the 2013 tournament, including that matchup between Wichita State and La Salle. Out of the 68 invited teams, there are 2278 possible matchups that could happen (all combinations of two teams, so it's 68*67, divided by two because we only want the matchups where the first team's ID is less than the second team's ID). So the submission would include 2278 data rows, and indeed includes the prediction for the potential matchup between Wichita State and La Salle:

1	ID	Pred	
2	2013_1103_1107	0.774055	#12 Akron with 77.4% chance to beat #15 Albany NY
3	2013_1103_1112	0.325755	#12 Akron with 32.6% chance to beat #6 Arizona
	...	...	
1038	2013_1211_1334	0.914999	#1 Gonzaga with 91.5% chance to beat #15 Pacific
1039	2013_1211_1338	0.557774	#1 Gonzaga with 55.8% chance to beat #8 Pittsburgh
1040	2013_1211_1355	0.894490	#1 Gonzaga with 89.4% chance to beat #13 S Dakota St
1041	2013_1211_1361	0.715163	#1 Gonzaga with 71.5% chance to beat #7 San Diego St
1042	2013_1211_1380	0.950888	#1 Gonzaga with 95.1% chance to beat #16 Southern Univ
1043	2013_1211_1387	0.654050	#1 Gonzaga with 65.4% chance to beat #4 St Louis
1044	2013_1211_1388	0.693812	#1 Gonzaga with 69.4% chance to beat #11 St Mary's CA
1045	2013_1211_1393	0.628855	#1 Gonzaga with 62.9% chance to beat #4 Syracuse
1046	2013_1211_1396	0.842221	#1 Gonzaga with 84.2% chance to beat #9 Temple
1047	2013_1211_1417	0.784193	#1 Gonzaga with 78.4% chance to beat #6 UCLA
1048	2013_1211_1424	0.757594	#1 Gonzaga with 75.8% chance to beat #5 UNLV
1049	2013_1211_1433	0.687924	#1 Gonzaga with 68.8% chance to beat #5 VA Commonwealth
1050	2013_1211_1434	0.838263	#1 Gonzaga with 83.8% chance to beat #14 Valparaiso
1051	2013_1211_1437	0.787186	#1 Gonzaga with 78.7% chance to beat #9 Villanova
1052	2013_1211_1443	0.957526	#1 Gonzaga with 95.8% chance to beat #16 WKU
1053	2013_1211_1455	0.746152	#1 Gonzaga with 74.6% chance to beat #9 Wichita St
1054	2013_1211_1458	0.585838	#1 Gonzaga with 58.6% chance to beat #5 Wisconsin
1055	2013_1217_1228	0.276911	#14 Harvard with 27.7% chance to beat #7 Illinois
	...	...	
1456	2013_1247_1437	0.450218	#13 La Salle with 45.0% chance to beat #9 Villanova
1457	2013_1247_1443	0.833082	#13 La Salle with 83.3% chance to beat #16 WKU
1458	2013_1247_1455	0.394212	#13 La Salle with 39.4% chance to beat #9 Wichita St
1459	2013_1247_1458	0.238477	#13 La Salle with 23.8% chance to beat #5 Wisconsin
1460	2013_1251_1254	0.317098	#16 Liberty with 31.7% chance to beat #16 Long Island
	...	...	...
2278	2013_1443_1458	0.059040	#16 WKU with 5.9% chance to beat #5 Wisconsin
2279	2013_1455_1458	0.324886	#9 Wichita St with 32.5% chance to beat #5 Wisconsin

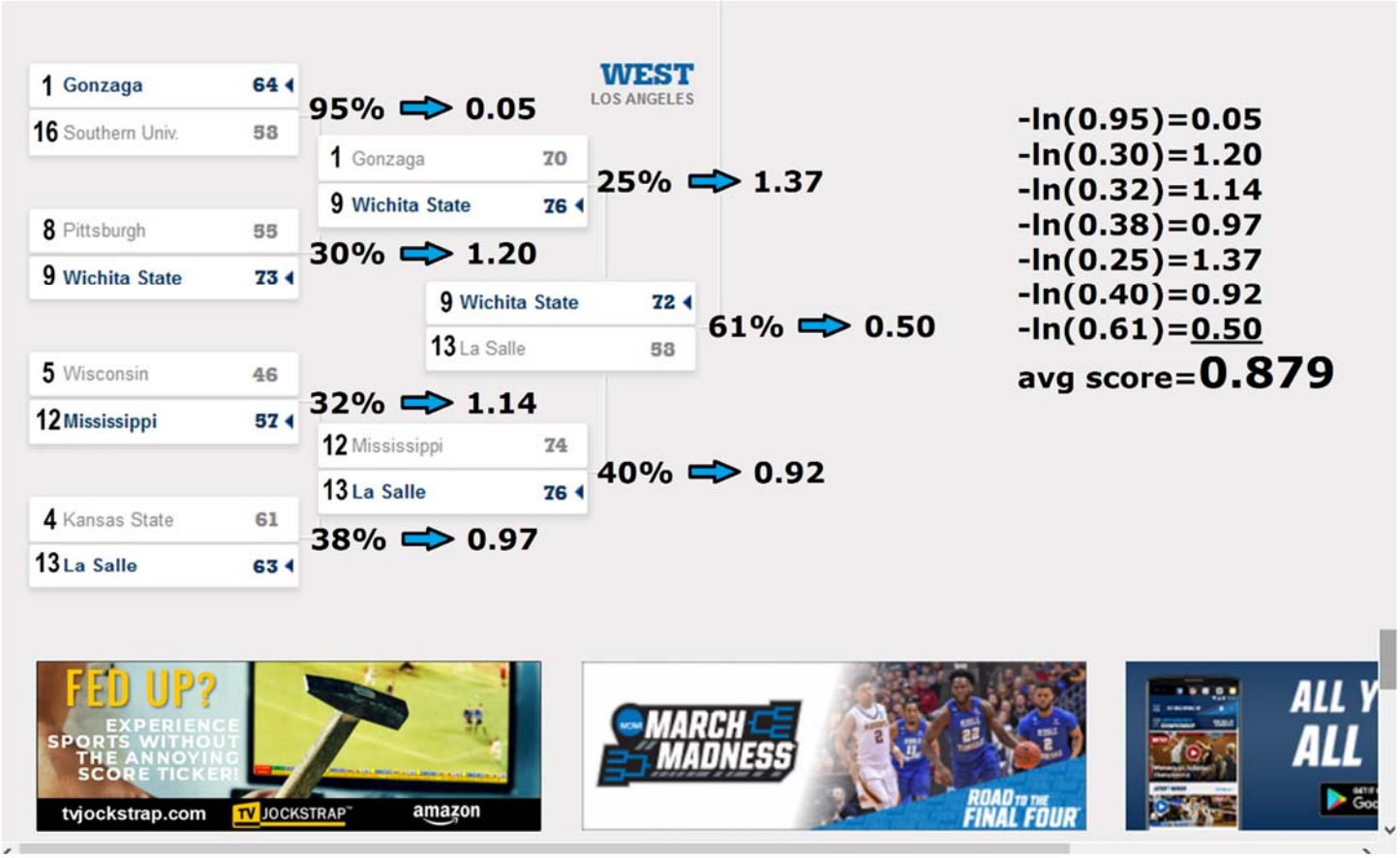
So in the Kaggle contest, the way we score a submission is to look at the games that were actually played - in this case we'll just look at the seven games in this half of the 2013 men's West Region again - and to note the predicted winning percentage from the perspective of the team that won each game.

For instance, #1 Gonzaga beat #16 Southern, and if we had made predictions based on the Ken Pomeroy national rankings, you can see from the above listing (row 1042) that we might have predicted about a 95% chance for Gonzaga to win. On the other hand, in the second round game, the Pomeroy rankings had Gonzaga ranked #4 nationally and Wichita State not even in the top-30, so you see that Wichita State would only be given about a 25% chance to win (row 1053). You can also see the 39% chance that La Salle has to beat Wichita State (row 1458), which translates to a 61% chance for Wichita State to beat La Salle.

So once the games are done, we will write down what our pre-tournament predicted winning percentage was for each game, from the perspective of the team that actually won the game. We will write 95% next to the Gonzaga-Southern U game, and 25% next to the Gonzaga-Wichita State game, and 61% next to the Wichita State-La Salle game, and so on.



And then for each prediction, we calculate the "log-loss" by taking the negative natural logarithm of each of these predictions. The worst performing predictions are the ones that give the biggest loss numbers.



So we can see that the single game with the worst prediction score, out of this part of the bracket, was that upset of #9 Wichita State over #1 Gonzaga, that only had a 25% predicted winning percentage for Wichita State, and so it got a loss score of 1.37. For the much more successful prediction of Gonzaga with a 95% chance to beat Southern, the loss score is only 0.05.

We take all of those log-loss scores and calculate their average, and that is your overall score on the leaderboard; in this case the average score across those seven games is 0.879. Generally speaking, this is an awful score for a big chunk of the bracket like this; even a simplistic 50% prediction for every game would have scored far better (a 50% prediction always scores 0.693), but of course that is not surprising, given the historic number of upsets in this part of the bracket.

It is worth emphasizing that even though there were upsets galore, and very few of the predicted teams advanced very far, we were still able to directly measure the accuracy of everyone's prediction for the games that really happened. It actually matters whether you thought Mississippi was better than La Salle, going in, or how much better you thought Wichita State was than La Salle. In the Kaggle contest, all games count equally, so there are 63 games you are scored against, and all of your other predictions are ignored, since they correspond to games that didn't end up happening.

It is also worth emphasizing that in the Kaggle contest, we don't only care who you think is predicted to win; we also care how confident those predictions are. As an example, let's look at the other three first round results, where #9 Wichita State and #12 Mississippi and #13 La Salle were all underdogs, based on the predictions from the Ken Pomeroy national rankings, and all three underdogs actually won.

We see that Wichita State was given a 30% chance to win its first round matchup - since Pittsburgh was actually considered by Pomeroy to be a top-ten team - whereas Mississippi had a 32% chance to win and La Salle had a 38% chance (Kansas State was considered by the Pomeroy rankings to be quite overrated as a #4 seed). So even though all three were upsets, they are scored differently on a game-by-game basis. The 30% prediction - for Wichita State to win - gets a worse score than the other two, because the prediction that Wichita State would lose was the most confident of the three.

Over the entire bracket, a universal prediction of 50% will be extraordinarily unlikely to be competitive. But the games where you make a more conservative prediction, closer to 50% than the majority of other competitors, could make a huge difference toward your score, even if you did agree with everyone else regarding the more straightforward question of who was the favorite.