

April 16, 2016

Project Specs

DA602 • Data Science in Python Web Framework
ANDREW HERNANDEZ

TOPIC	DESCRIPTION
The Data	<p>For the years 1970 through 2016, this wikipedia page lists the final four participants, along with their tournament seed. (It also lists the 2 teams that make the finals, as well as the winner)</p> <p>https://en.wikipedia.org/wiki/List_of_NCAA_men%27s_Division_I_basketball_tournament_Final_Four_participants</p>
The Stats	<p>If the size of your pool was 2, then simply selecting the top seeds might quite easily win the pool. As the number of entrees in the pool increases, the more upsets it will require to win the pool. I would like to show, therefore:</p> <ol style="list-style-type: none">1. The historical sums (chart) of final four participant team seeds in the final 42. Based upon the number of pool entries, by random sampling, I would like to show certain pick distributions to be more probable winners than others.
The Architecture	<p>I would like to:</p> <ul style="list-style-type: none">• Parse the final four data from the wikipedia page above. This will be done via a script, and not part of any webapp.• Write the final four and seeds information to a MySQL DB (also via a script)• Render this information via the web on a free Python hosting site, such as https://www.pythonanywhere.com• I'm leaning towards either Web2Py, or not using a framework at all, still deciding on this.• I'm thinking of a stacked bar chart such as this one: http://canvasjs.com/docs/charts/integration/jquery/chart-types/jquery-stacked-column-chart/ to show the overall sum of the seeds that historically makes the final 4.• Show the historical odds of different seed combinations making the final four, ie:<ul style="list-style-type: none">◦ What are the odds of no seeds below 3?◦ What are the odds of all seeds below 3?◦ What are the odds of a see between 6 and 8 making the final four?◦ What are the odds of 2 such teams making the final four?◦ Other (hopefully) interesting questions.• The target data frame should be simple, with the following core columns:<ul style="list-style-type: none">◦ YEAR◦ WEST_TEAM◦ WEST_SEED◦ EAST_TEAM◦ EAST_SEED◦ MIDWEST_TEAM◦ MIDWEST_SEED◦ WEST_TEAM◦ WEST SEED

The Web	<ul style="list-style-type: none">○ FINALS_WINNER○ FINALS_WINNER_SEED○ FINALS_LOSER○ FINALS_LOSER_SEED <p>I would like to show some interesting results with this data on the web, and would also like to use some type of javascript/charting technology to represent some of the findings graphically.</p>