

March Madness Bracket Picker

By Weston Kunz

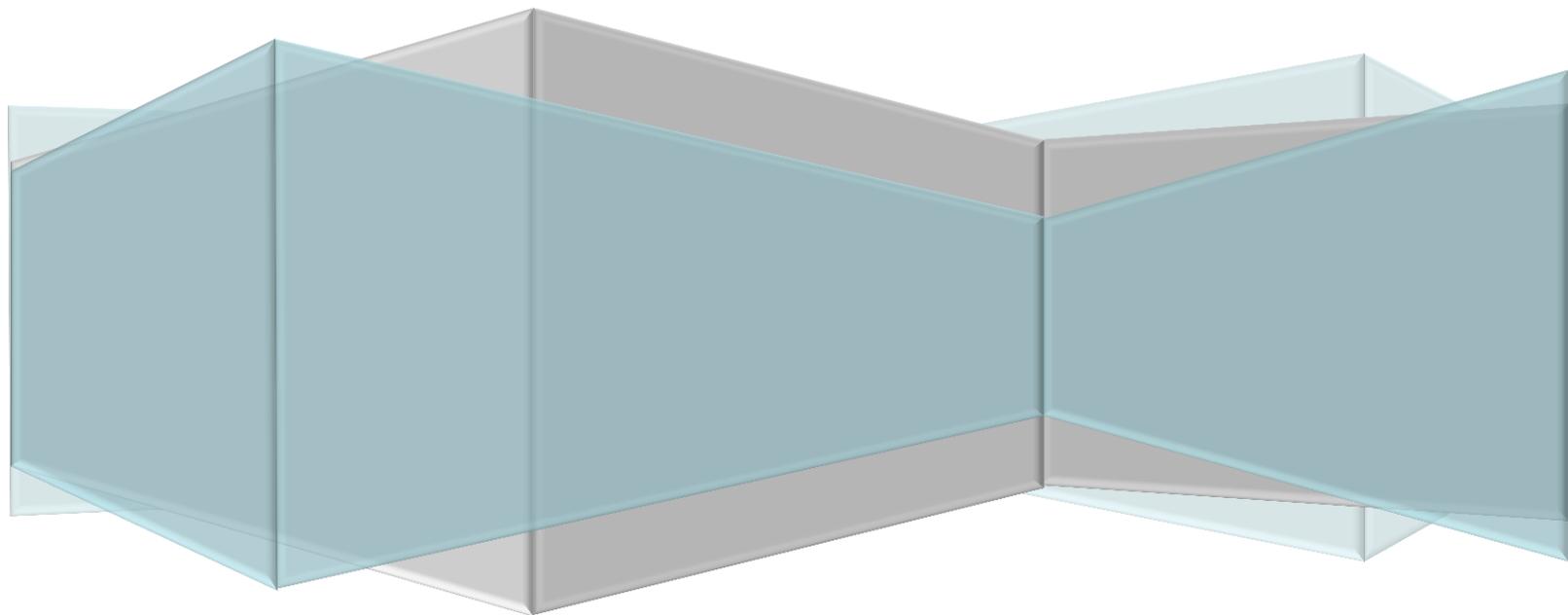


Table of Contents

Executive Summary.....	2
The Project.....	3
Original Plan.....	3
Upsets	3
Implementation	3
Class Modules.....	3
Setting up the data	4
Using collections	4
Ribbon.....	5
Team Stats Form.....	5
Bracket.....	6
Introduction Page.....	7
Lessons Learned	8
Assistance.....	8

Executive Summary

The NCAA basketball tournament is one of the most hyped sporting events of the year. Millions of people around the nation participate in March Madness bracket competition, during which time they try to predict the outcome of every game during the tournament. Although the chances of predicting every game correctly are slim to none, that doesn't stop sports fans from taking pride in that fact that their bracket is not as bad as that of their peers. Unfortunately, the hype that takes place in the middle of March as the tournament begins quickly dies down as their predictions prove to be incredibly incorrect.

In attempt to bring some order to the "madness" I created the March Madness Bracket Picker. This workbook reads stats for all 344 NCAA Men's division I basketball teams and chooses the outcome of each game according to the stats. The predictions are scored against the actual outcome of the 2012 Men's basketball tournament. The scoring system used by the bracket is the same scoring system that ESPN uses for its bracket challenge. Along with a bracket generator I have included a form that displays the stats for each team. This form is helpful for the careful planner who wants to see the stats himself/herself and then make his/her own predictions. More than anything, this workbook helps the user understand which stats help make the best predictions.

The Project

Original Plan

The original scope of this project was to create a workbook containing a button that would gather data for every NCAA basketball team. That data would then be used to populate a form providing an organized look at the stats for each individual team. Also, the form was to contain the logo for each team. Along with creating a form, a variable called the “determinant variable” was going to be created for each team. The variable would be calculated by giving certain weights to each stat available. The determinant factor would then be used to generate a March Madness tournament bracket. The team with the higher determinant factor in each game would be predicted the winner.

Upsets

It wasn't long before a few of my dreams were shattered. After a few futile attempts of pulling in the logo for each team from the Internet, I came to the realization that it wasn't possible. In order to include the image I would have to create a file containing all of the team logos and then the form could access the images from there. This was a problem but not necessarily a deal breaker.

I spent hours creating a web scraper that accessed Statsheet.com/mcb and gathered data for every team. First I had to scrape one page containing the name of every team. Because often times the team name did not match the name the URL used, I had to loop through the hyperlinks on the page and extract the names the site uses for each team. My scraper took approximately 30 minutes to run and hit nearly 700 different webpages. After experimenting with my scraper several times and having gather 4 years of basketball data I found I had been blacklisted from StatSheet.com. Or that is to say, I got BYU blacklisted from StatSheet.com. Just a day or two earlier professor Allen had mentioned a time when he got another university blacklisted for doing the same thing, and when he confronted the website admin two years later the admin said that he would have given professor Allen data had he known it was for academic purposes. With this story fresh on my mind I contacted the site admin of StatSheet.com. He rejected my suggestion of giving me or allowing me to scrape more data despite the fact that I was using it for an academic purpose. A few emails later and we settled on BYU being removed from their blacklist if I agreed not to scrape their site again.

It was great to be removed from the Blacklist but it threw a wrench into my whole project. No longer would I be able to create a button to scrape the data for future years. This also meant that I would no longer be able to scrape StatSheet.com for logos for every basketball team.

Implementation

Class Modules

As I began working on this project I quickly realized that I wanted to use object oriented programming. Although we didn't cover class modules in class I quickly figured out how to

create one and create the procedures to manipulate the properties of each class. The project revolves around to objects/classes: team and conference. Each conference has an ID and a conference name and each team has an ID, a team name, a conference, and many more stats.

Setting up the data

The data for this project was scraped from StatSheet.com/mcb during one of my trial runs of my scraper. Because I was no longer going to be able to create a button to gather data for each future year (see “Upsets” section above) I decided I needed to do something to make my results stand out a little bit more. I decided I would combine concepts I was learning in another one of my classes this semester: Data Mining. I took the four years of data that I had scraped during my trial runs. This data consisted of basketball stats of every NCAA basketball team from 2009 – 2012. I used 2009-2011 to create a model. I then applied the 2012 stats to the model and created a predictive variable called “predicted rounds”. Essentially what this variable says is that based on past basketball tournament outcomes, this team is predicted to make it this far in the tournament in 2012. The 2012 Stats, including the “predicted rounds” variable for each team were placed on a worksheet in this project (See figure 1).

The following basketball stats were scraped from the www.StatSheets.com/mcb on March 15, 2012.

ID	Team Name	Conference	Wins	Losses	Conference Rank	RPI Rank	SOS Rank	Field Goal Percentage	Free Throw Percentage	Three Point Percentage
1	Syracuse	Big East	32	2	1	1	28	46.5	69.2	34.1
2	Kentucky	SEC	33	2	1	2	38	48.5	72	37
3	Michigan State	Big Ten	27	7	1	3	1	47.7	69.7	36.6
4	North Carolina	ACC	29	5	1	4	6	46.1	68.2	33.9
5	Duke	ACC	27	6	2	5	3	45.8	70.2	37.6
6	Kansas	Big 12	27	6	2	6	11	48.5	69.6	35.8
7	Ohio State	Big Ten	27	7	1	7	10	48.3	69.8	32.6
8	Marquette	Big East	26	7	2	8	17	45.9	72.1	34.1
9	Baylor	Big 12	28	7	3	9	11	47	75	38.1
10	Florida State	ACC	24	9	3	10	4	45.8	70.3	35.3
11	Missouri	Big 12	30	4	1	11	96	50.4	76.6	39.5
12	Georgetown	Big East	23	8	4	12	13	46.3	69.7	34.7
13	Louisville	Big East	27	9	3	10	7	42.4	68.8	31
14	Indiana	Big Ten	25	8	3	14	32	48.7	76.2	43.3
15	Michigan	Big Ten	24	9	5	15	16	45.6	72	35.2
16	Wichita State	Missouri Valley	26	6	2	20	62	48.1	75	36.8
17	Memphis	C-USA	26	8	1	17	25	49.4	72	36.9
18	Temple	A-10	24	7	2	18	59	47.2	71.8	40.2
19	Unlv	Mountain West	26	8	3	19	49	46.2	67.9	37
20	Vanderbilt	SEC	25	10	2	16	5	46.1	69.8	39.1
21	Southern Miss	C-USA	25	9	2	28	59	41.5	73.6	35.8
22	Wisconsin	Big Ten	25	9	4	21	18	42.5	74.6	36.5

Figure 1 – Team Stats scraped from StatSheets.com

Using collections

Each row was used to create a team object and immediately placed into a public collection called “Teams”. By placing the team objects into the collection I was easily able to cycle

through the collection throughout the course of the project. I created another collection of conference objects that was used in the creation of drop down menus in the form.

Ribbon

Using Professor Allen's RibbonWizard workbook I created a "March Madness" tab on the excel ribbon for this project. The tab contains two sets of buttons (See Figure 2).

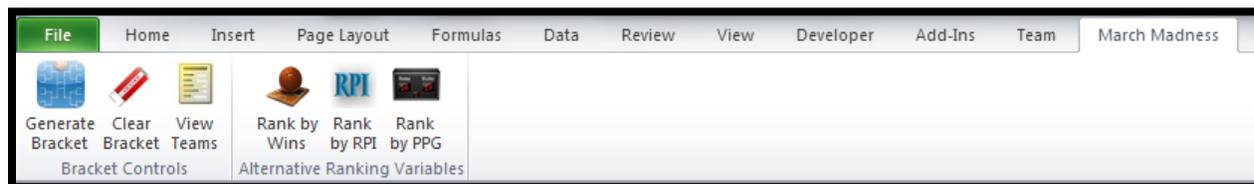


Figure 2 – Contents of the "March Madness" tab in the workbook ribbon

The first set of buttons is called "Bracket Controls". This set contains a button to generate the outcomes of each game in the bracket using the default predictive variable: "predicted rounds". Another button on this ribbon is set up to clear the results of the bracket. This includes removing team names, scores, and highlighting produced from the bracket generation. A third button in this section brings up the form that contains the statistical information for every team.

The second set of buttons is called "Alternative Ranking Variables". The buttons in this set provide other ways to generate the tournament bracket and predict a winner. This set gives you the option of predicting based on the number of wins each team has, their RPI rank, or the number of points each team scored per game during the season. These are alternative methods that people generally use to predict who they think will win each game in the tournament. It doesn't take long to notice that although some of these variables are affective, they don't compare with basing your assumptions off of trends from past years.

Team Stats Form

In generating the team stats form it loads a conference comb-box containing the names of every conference in Division I college basketball. When a conference is selected the school combo-box is loaded based off the name of the conference. The team collection is looped through and every team where their conference property matches the selected conference gets added to the school combo-box. When a team/school is selected another loop through the collection is made. This loop finds the team that matches the selection and loads all of the stats into the form. This form is great for the fan that wants to either predict the bracket himself/herself or just take a closer look at why the buttons generated the brackets in the way they did (See figure 3).

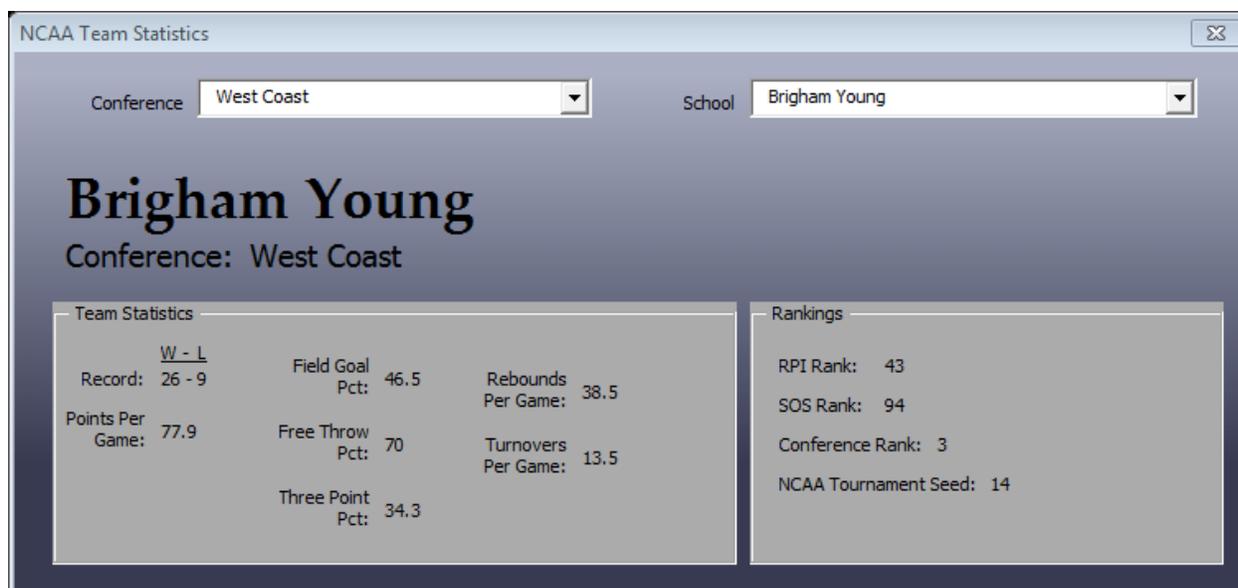


Figure 3 – Team Stats form displaying the 2011-2012 stats for the BYU Cougars Men’s basketball team

Bracket

The Tournament bracket is the center of attention in this workbook. Everything else in the workbook works together to make the bracket as efficient as possible. When the “Generate Bracket” button is pushed the bracket is generated 1 round at a time with a second delay between each round. This gives the view a chance to see how the bracket is actually progressing. The bracket is constantly being compared against the actual outcome of the tournament (located on the “Actual Bracket” worksheet). If a prediction is incorrect the cell is highlighted in red giving the user a quick and easy way to view missed predictions. The winner of the bracket is displayed in a large font in blue, making sure there is no question about the final prediction. As each round is generated a score is updated above the bracket displaying the current number of points achieved by the prediction. As stated above, the bracket is quickly and easily restored to its previous state by clicking on the “Clear Results” button in the “March Madness” tab of the ribbon (See figure 5).

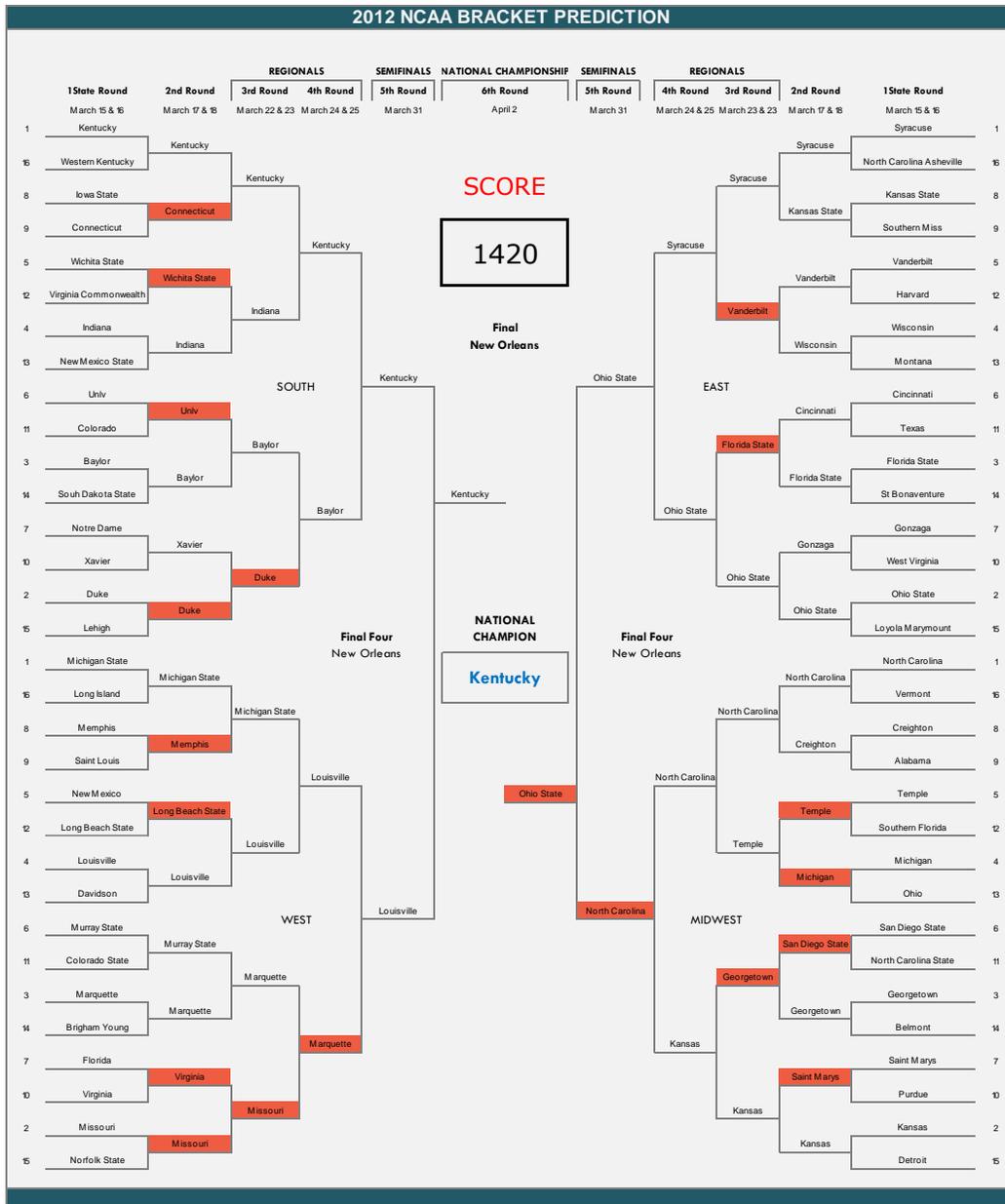


Figure 5 – Bracket generated using the “Generate Bracket” button in the “March Madness” Ribbon tab

Introduction Page

In light that there is a chance that not every user will have the opportunity to review this document before using the workbook I have included an introduction page giving an in-depth outline of what the workbook does. There is nothing worse than getting a tool and having no idea how to use it (See figure 6).

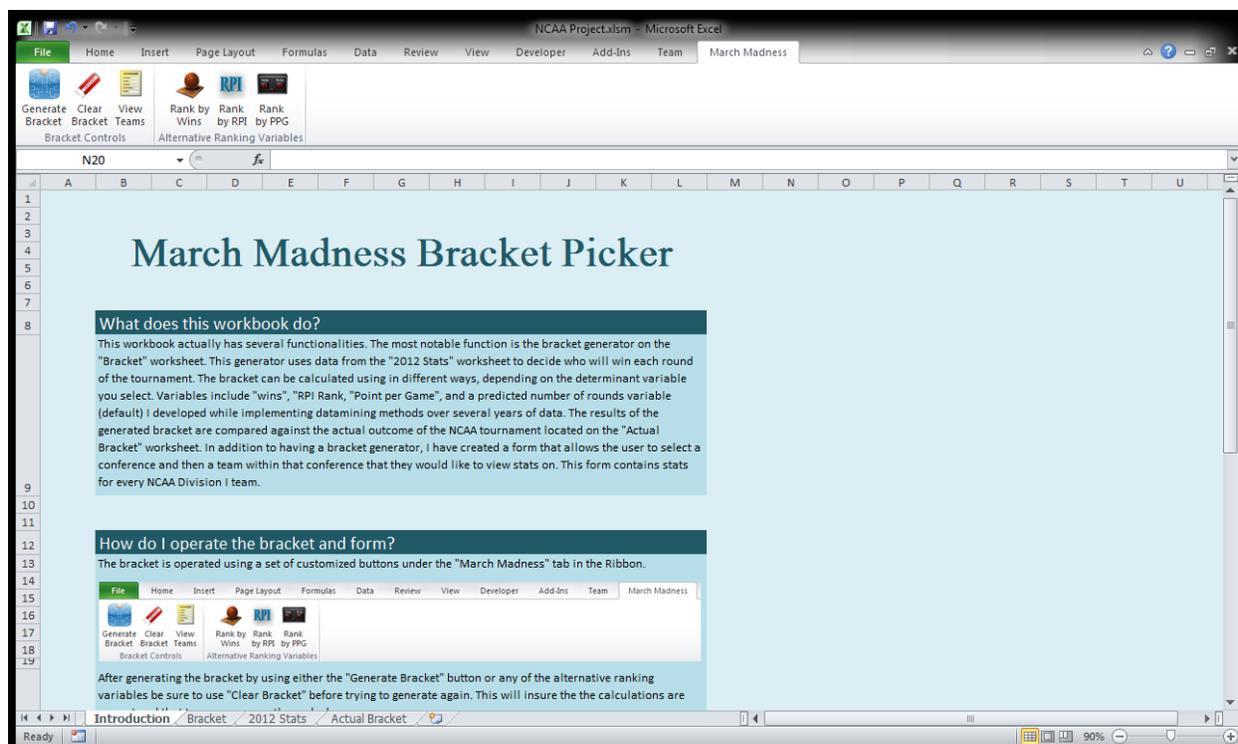


Figure 6 – Introduction worksheet with descriptions of the functionality of the workbook.

Lessons Learned

There were several important things I picked up from fighting with this project. I learned how to and how not to scrape large amounts of data. Perhaps I should have questioned if I should be scraping data after we had our class about checking the robot.txt file on websites to see if scraping is allowed. I practiced project oriented programming in VBA for the first time and found it to be rather rewarding when dealing with the same data over and over. Before working on this project my understand of collection, maps, and dictionaries had always been a little off, but by immersing myself into understanding these objects I was able to successfully choose the one that best suited my needs. I now find myself constantly watching for situations where I can automate a process using VBA.

Assistance

Besides one or two quick questions to professor Allen, I did not receive assistance on this project from any other individual.