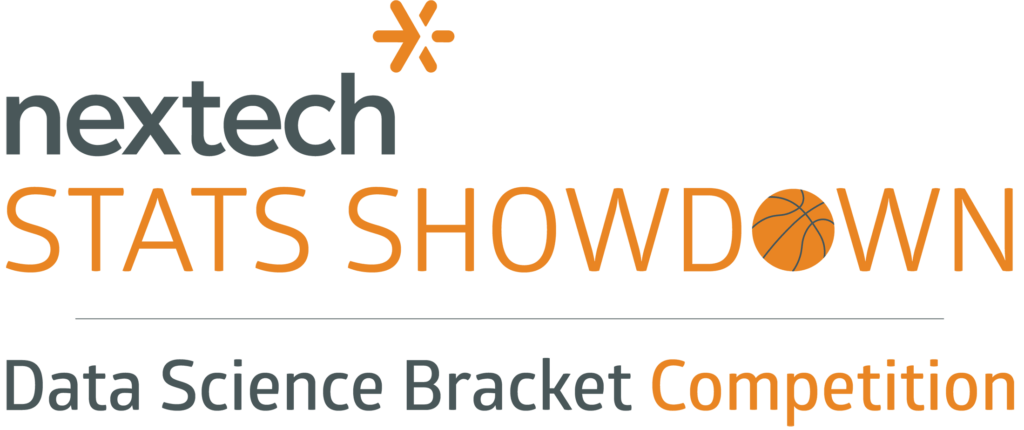
****

**Rules & FAQ**

**This competition is modeled after The Iowa Center for Interdisciplinary Training’s program led by Rick Spellerberg.**

**Table of Contents**

[Competition Overview](#_gjdgxs)

[Rules for the Competition](#_30j0zll)

[Stats Showdown Scoring](#_3znysh7)

[Stats Showdown Data Set](#_1t3h5sf)

[Algorithm and Top Ten Teams Submission Process](#_4d34og8)

[2025 Competition Dates](#_17dp8vu)

[Prizes](#_26in1rg)

[Example Algorithms](#_35nkun2)

[Ideas for Analyzing Data in Spreadsheets](#_1egqt2p)

[Ideas for Developing Better Algorithms](#_3ygebqi)

### Competition Overview

The Nextech Stats Showdown Basketball Data Science Bracket Competition is a competition for teams of 6th-12th grade students (to be called “Student Groups” for the remainder of this document so as not to cause confusion with the basketball teams) to introduce them to the exciting world of data science and data analytics through March Madness and the NCAA tournament.

The goal of this competition is that students will answer the question ***“What team in the NCAA Tournament is going to win the most games throughout March Madness?”*** Teams will be tasked with using past tournament teams’ data from the past year to create an algorithm that most accurately predicts the game winners throughout the tournament. Teams will then use that algorithm and apply it to the 2025 data after Selection Sunday (March 17) and submit their Top Ten list for 2025.

Prior to the announcement of the 2025 NCAA Basketball Tournament field, teams will submit an algorithm that will choose the ten teams from the tournament field most likely to win games. After Selection Sunday, Student Groups will use their algorithm to submit their team’s top ten list. Who are the top ten teams (numbered from 10 - 1, 10 being the most likely to win the most games and 1 being the least) that are going to win games during March Madness as determined by their number crunching?

Each time a team in the list wins a tournament game, points are earned equivalent to the team’s assigned number in the list. So, for example, if the team in the number 10 spot, won the tournament, they would receive 60 points. If another’s team algorithm, put the winning team in the #1 spot, they would receive 6 points. ​​

**Cinderella Teams**

A team is considered a Cinderella team if they were seeded 9-16 in the tournament, so if a team in that list wins a game, that team will earn an additional 5 bonus points per round. For a first round win, a Cinderella team will earn 5 bonus points, for a second round win the bonus earned will be 10 points, etc.

For example, if a team in the list that is seeded 9-16 wins a game in the tournament, that team will earn additional five bonus points as a “Cinderella” team. For a first round win, a Cinderella team will earn five bonus points, for a second round win the bonus earned will be ten points, fifteen for the third round, etc.

Using the example above, if the winning team was a Cinderella team and was in the number 10 spot on a submission, the team would receive ((10\*6)+(5+10+15+20+25+30)) 165 points. If the same team was placed in the number one spot on the submission, the team would receive ((1\*6)+(5+10+15+20+25+30)) 111 points.

### Rules for the Competition

1. Each Student Group must have an adult sponsor that registers the team. This adult can be a teacher, parent or guardian, club advisor, etc. The adult sponsor will be responsible for making sure the student groups comply with the rules and meet all the requirements and deadlines for the algorithm submission.
2. Student Groups can consist of 2-4 students, from grades 6-12. Student Groups can have mixed grades.
3. Each Student Group can only submit one algorithm.
4. The algorithm will be initially submitted to the sponsor before Selection Sunday. After Selection Sunday, sponsors will be provided the submission form (a google form) along with the current year data set. Student Groups will use this form to submit their algorithm and their selections (10-1, numbered in order from highest to lowest) no later than midnight on March 18th.
5. A template will be provided to the sponsors for algorithm submission.
6. In terms of guidance, sponsors can teach their students any concepts or train them on any tools that they can use that may help them in the development of algorithms. For example, sponsors can coach the students on how to use tools found in Google Sheets, Microsoft Excel, or statistical software such as R and SAS or teach them statistical concepts that can be useful such as calculating z-scores or correlations. ***Students are allowed to manipulate or create any additional data with the original data provided. They are not allowed to pull additional data from other sources.***
7. Sponsors are not allowed to provide examples or assist in the development of algorithms. Sponsors can answer questions students may have as they develop algorithms. This competition will rely on the honor system. Some example algorithms are provided later in this document and are not eligible for submission.

### 

### Stats Showdown Scoring

Each team will submit a list of the ten teams from the tournament field most likely to win games, numbered 10, 9, 8, 7… 2, 1.

|  | **Team** | **Rank** |
| --- | --- | --- |
| Most likely to win the most games | College A | 10 |
|  | College B | 9 |
|  | College C | 8 |
|  | College D | 7 |
|  | College E | 6 |
|  | College F | 5 |
|  | College G | 4 |
|  | College H | 3 |
|  | College I | 2 |
| Least likely (of these 10 teams) to win the most games | College J | 1 |

Using this example, if College C were to win the tournament (win all 6 games, the most a team can win in the tournament), they would receive 48 points for that team (6 games \* 8 points per game = 48 total points).

However, let’s say that College I actually wins the tournament instead. The submission would now only earn 12 points for the championship (6 games \* 2 points per game).

Let’s go a step further, let’s say that College I was actually seeded 9 or higher in the tournament. This would make them a Cinderella team. Cinderella teams earn an additional 5 bonus points per round where each round is 5 *additional points (*ie, round 1 = 5 additional points, round 2 = 10 additional points, etc.). So, now, College I would receive (6 games \* 2 points per game) + (5+10+15+20+25+30 for all their accumulated Cinderella points) for a total of 117 points.

**Maximum Possible Points**

This table shows you that the maximum number of points a team can earn with no Cinderella teams and perfect team placement in terms of how many games they win is 225 points.

If you place every team perfectly in the algorithm rating AND all 10 of your teams are Cinderella teams, the best possible score is 655 points.

Note: Play-in games (First 4) don’t count in scoring.

| **Algorithm Rating** | **Tournament Games Won** | **Stats Showdown Points Earned (maximum with no Cinderella teams)** | **Cinderella Extra Points (best case)** | **Stats Showdown Points Earned (all Cinderella teams)** |
| --- | --- | --- | --- | --- |
| 10 | 6 | 60 | 105 | 165 |
| 9 | 5 | 45 | 75 | 120 |
| 8 | 4 | 32 | 50 | 82 |
| 7 | 4 | 28 | 50 | 78 |
| 6 | 3 | 18 | 30 | 48 |
| 5 | 3 | 15 | 30 | 45 |
| 4 | 3 | 12 | 30 | 42 |
| 3 | 3 | 9 | 30 | 39 |
| 2 | 2 | 4 | 15 | 19 |
| 1 | 2 | 2 | 15 | 17 |
| **Total Points Possible:** |  | **225** |  | **655** |

### 

### 

### 

### Stats Showdown Data Set

There is a wide variety of data points that exist in the full data set for each year. For a full list of the column data shared and an explanation of each, please visit this [Explanation of Stats and Data in Data Sets document](https://docs.google.com/document/d/16vdBcRjVD8r-t1Ph8zQqgqTdUXB8qxcTDaX2Af41Tbs/edit?usp=sharing).

**What are you allowed to do with the provided data?**

* You can sort, filter, manipulate the data in any way that you find helpful. This includes things like writing functions, using computer functions, graphs, calculating averages, etc.
* You can create equations with multiple data points in them and use the results to sort/select teams
* You cannot add any stats that are not shared in the dataset.

### Algorithm and Top Ten Teams Submission Process

1. On March 14th, 2025 before the tournament field is announced, sponsors will collect all the algorithms submitted.
2. Once the tournament field is announced Sunday March 16th, 2025, the 2025 data set will be added to the Stats Google Sheets (both Men’s and Women’s) and made available to the sponsors as soon as possible. The goal is by the end of the day on March 16, 2025. If this date changes, the submission deadline will change accordingly.
3. Once in their possession, the sponsors provide the 2025 data set to their student groups so they can evaluate and run their algorithms and determine their team lists for eventual submission.
4. For accuracy purposes, sponsors can either check themselves or have teams exchange their algorithms to check for accuracy. Accuracy means that the basketball team selections must be able to be replicated exactly by non-team members through following their submitted algorithm.  **We advise having student groups exchange their algorithms with other teams as a component of the learning experience.**
5. Student Groups will submit their algorithm and their basketball team selections on a google doc shared with the 2025 data set. The student groups (or the sponsor) can complete the submission. All submissions must be received by midnight on March 18, 2025.
6. Sponsors will be provided Student Group scoring updates during the competition to let their teams know where they stand as well at the end of the competition.
7. Winners will be announced no later than April 18, 2025.

### 

### 2025 Competition Dates

| February 1 (or after) | The dataset is released to sponsors who are already registered, and student groups can start creating their algorithm. If sponsors register after February 1, they will receive the data set and instructions within 2 business days. |
| --- | --- |
| Friday, March 14, 2025 | Algorithms due to sponsor by end of the day. |
| Sunday, March 16, 2025 | Selection Sunday |
| No later than Tuesday, March 18, 2025 | Sponsors are provided the 2025 data sheet. |
| March 19-20, 2025 | Teams have 48 hours after the Selection Sunday team data is distributed to apply their algorithm to the data, select their top 10 teams and submit their selections (and algorithm used) to Nextech. |
| Throughout Tournament | *Teams will be provided updates on Stats Showdown Data Science Bracket Competition results.* |
| March 20-23, 2025 | Rounds 1 and 2 of the NCAA Tournament |
| March 27-30, 2025 | Rounds 3 and 4 of the NCAA Tournament (Sweet 16 & Elite 8) |
| Saturday, April 5, 2025 | The Final Four |
| Monday, April 7, 2025 | The Championship Game |
| Friday, April 18, 2025 | Winners notified on or before date; prizes distributed. |

### 

### Prizes

The ***winning team*** will win a traveling Nextech Stats Showdown trophy to display in their school, a pizza party, and bragging rights for the year.

The pizza party will be determined based on where the team is housed -- for example, if an after school club or via multiple classes, the pizza party will be provided for the team winners.

If the Stats Showdown was completed in class, the pizza party will be provided for the winning team’s class.

The ***school with the highest team scores average*** will win a pizza party for the team members from their school.

All winners will also receive certificates.

### 

### Example Algorithms

Algorithms can be as simple as “Sort the current year tournament teams’ data set on ‘Season Wins’ and rank the top team and #10, second highest team with #9, etc for the top 10 winningest regular seasons.” OR as complex as creating data models and equations using math and computer programming software such as MATLAB or similar.

***Note:*** These examples are intended to help start helping develop ideas of what is possible and to get the creativity flowing for students interested in participating in this competition. Feel free to use these examples to help you come up with ideas, but we would not recommend using these exactly as written (as they are intentionally not using the best data sources to pick the top teams).

**Example 1:**

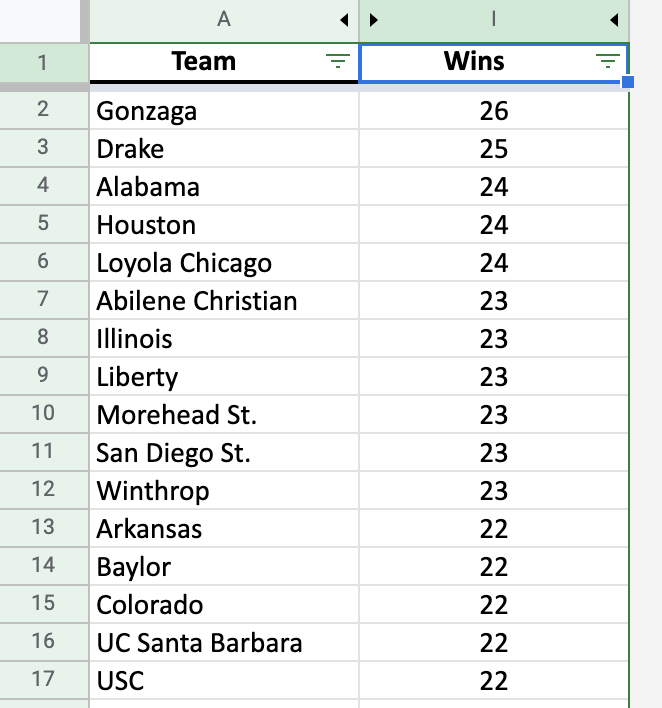
**Proposed Algorithm**: Order the teams in the tournament by largest to smallest number of regular season wins. Assign the team with the highest number of wins the number ten, the second highest the number 9 and continue this process until you get to the tenth highest value and assign that team the number 1. If in a step of the process there are two or more teams with the same number of regular season wins, assign the values using alphabetical order via the school’s name.

This example is using the 2021 data set to generate the algorithm and then testing all other year’s data sets against the algorithm to determine strength.

**Testing the Algorithm’s Performance:**

Now we must test this algorithm for performance. We will do this by simply sorting the google sheets document on Column I: “Wins.”

Let’s first consider the 2021 tournament results. We have hidden all the columns that we do not need for our algorithm. Then, we filter the data and sort Z → A for the wins column.



The teams starred were seeded 9-16 making them eligible for Cinderella bonus points (these points are put in parentheses in the points earned column).

| **Team** | **Algorithm Rating** | **Tournament Games Won** | **Stats Showdown Points Earned** | **Conference** |
| --- | --- | --- | --- | --- |
| Gonzaga | 10 | 5 | 50 | West |
| Drake\* | 9 | 0 | 0 | West |
| Alabama | 8 | 2 | 16 | Midwest |
| Houston | 7 | 4 | 28 | East |
| Loyola Chicago | 6 | 2 | 12 | Midwest |
| Abilene Christian\* | 5 | 1 | 5(+5) =10 | East |
| Illinois | 4 | 1 | 4 | Midwest |
| Liberty\* | 3 | 0 | 0 | Midwest |
| Morehead St.\* | 2 | 0 | 0 | Midwest |
| San Diego St. | 1 | 0 | 0 | Midwest |

**Total Points Earned = 120 Maximum Points Possible = 655**

Now if you test the algorithm on the 2013 through 2019 tournaments, the algorithm will perform at the following levels:

**2013 = 88 points 2014 = 110 points 2015 = 165 points**

**2016 = 116 points 2017 = 58 points 2018 = 173 points 2019 = 115 points**

The average points earned for this algorithm over the five tournaments is 117.86. Considering a good average will be in the 170+ range, this algorithm appears to be a weak performer. There is one year, 2018, where the algorithm performs much better than the others, but this is offset by seven other years where the algorithm performs poorly. Analysis of the algorithm is now important. If you look at the results produced by the algorithm for the 2016 tournament and focus on the regions, you will notice that no team in the South Region was selected, this is a potential problem. At no point in the tournament will this algorithm earn points from teams in the South region for the 2021 tournament. This may be our problem.

**Example 2**

Based on our testing of the original proposed algorithm, there are 2 things we would like to attempt in this second iteration:

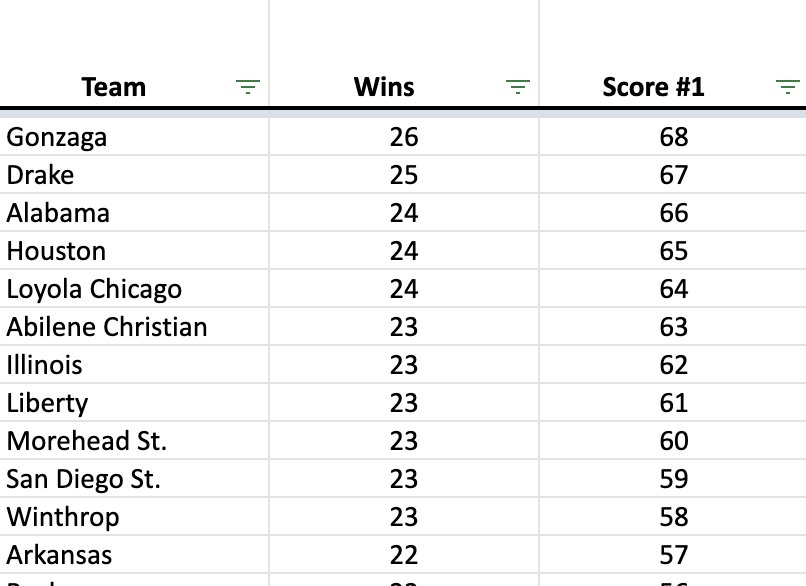
1. Be aware of which Region each of the teams is in
2. Add a second data point to rank the teams on

**Plan for next iteration of algorithm:**

1. Still use the regular season wins.
2. Add, “Free Throw Percentage” as another data point.
3. Make sure the final list has at least 2 teams from each region.

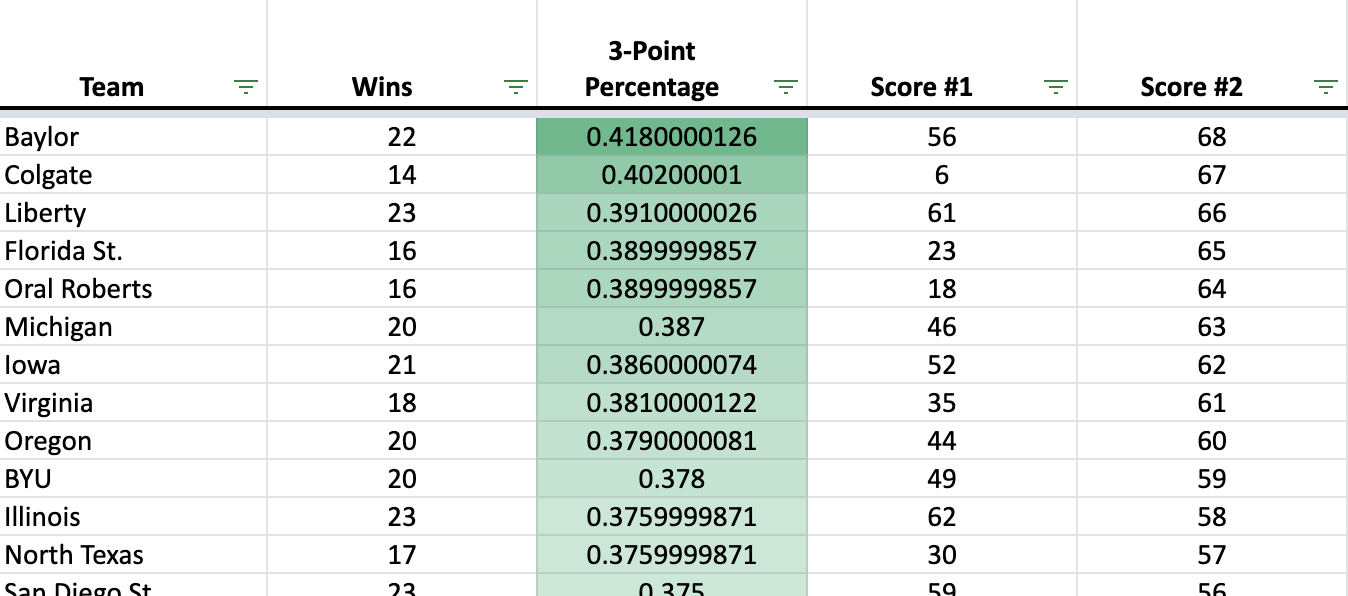
**Adjusted Algorithm**:

1. First, make sure the teams are sorted in alphabetical order. Then, order the teams in the tournament by largest to smallest number of regular season wins. Create a “Score #1” column and assign the team with the highest number of wins the number 68, the second highest the number 67 and continue this process until you get to the lowest value and assign that team the number 1. If in a step of the process there are two or more teams with the same number of regular season wins, assign the values using alphabetical order via the school’s name.



Demo of Step #1 for 2021 data

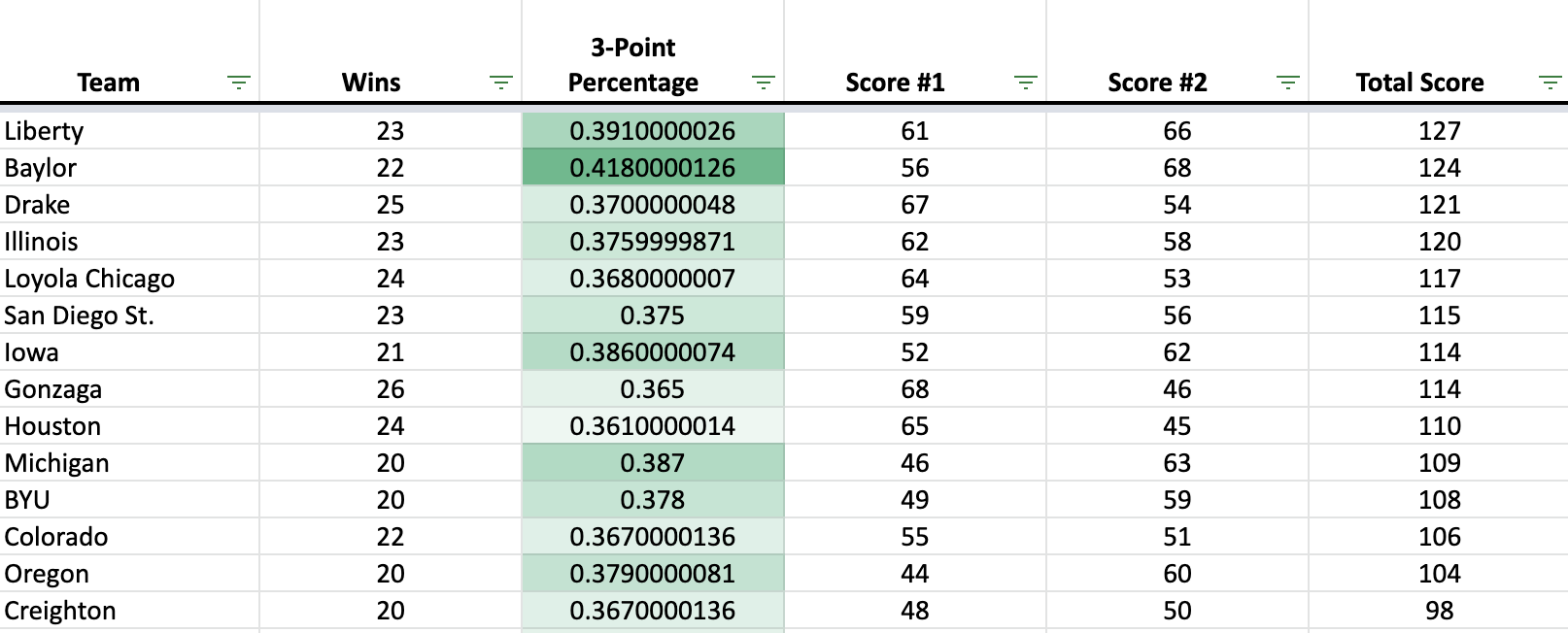
1. Next, sort the teams in alphabetical order again to reset that order. Then, order the teams in the tournament by largest to smallest 3-Point Percentages. Create a “Score #2” column and assign the team with the highest 3-point percentage the number 64, the second highest the number 63 and continue this process until you get to the lowest value and assign that team the number 1. If in a step of the process there are two or more teams with the same number of regular season wins, assign the values using alphabetical order via the school’s name.



Demo of Step #2 for 2021 data

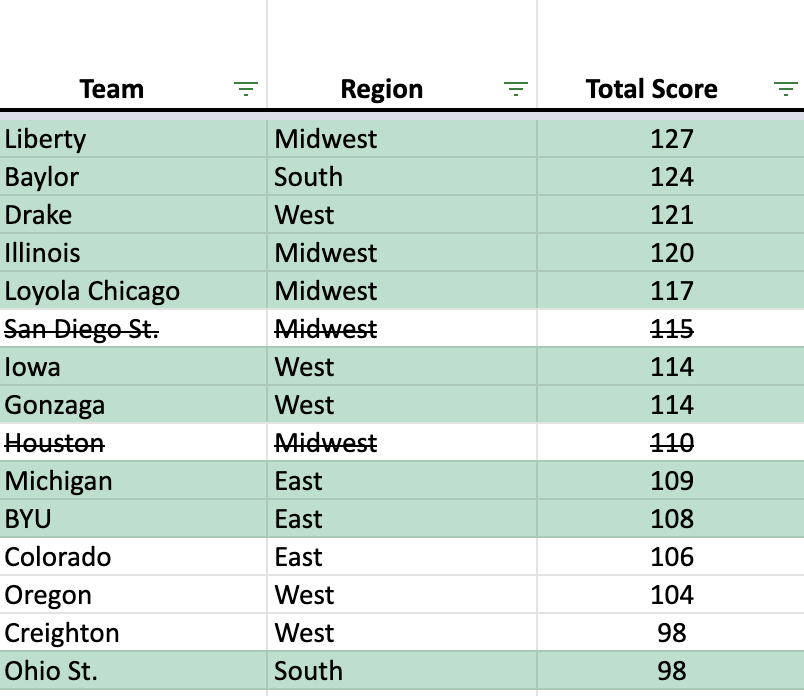
*(note: A color scale for conditional formatting was used for the 3-Point Percentage column when initially analyzing the data -* [*this YouTube video*](https://www.youtube.com/watch?v=oCxvm4Miz1w) *explains what it is and how to use it)*

1. Next, create a 3rd “Total Score” column to add up the Score #1 (from regular season wins) and Score #2 (from 3-Point Percentages). When you have the score for all 68 teams, sort the teams in alphabetical order again to reset that order and then sort the “Total Score” column from highest to lowest.



Demo of Step #3 for 2021 data

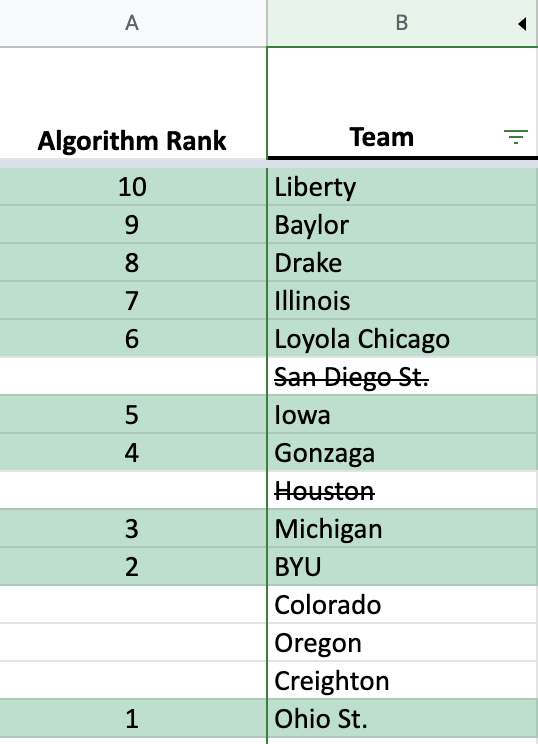
1. Make sure there are at least 2 teams from each Region in the top 10 list. If there are not, remove the lowest ranked team from the Region that has the most teams present in the top 10 list and add the next highest team from the Region that does not have 2 teams listed. Do this until there are at least 2 teams from each Region in the top 10 list.



Demo of Step #4 for 2021 data

(Teams highlighted in green are the ones selected for our Top 10 list, teams crossed out are from Regions that had the most representation and got removed from the original Top 10 list)

1. Assign each of the Top 10 teams (ordered by “Total Score” and accounting for Regional representation) in order from 10, 9, 8, etc until you get to 1.



Demo of Step #5 for 2021 data

**Testing the Algorithm’s Performance:**

Now we must test this algorithm for performance.

If we test the adjusted algorithm that takes 3-Point Percentages and Regional assignments of the teams into consideration for the 2021 tournament, we end up with the following results. *Note:* The teams starred were seeded 9-16 making them eligible for Cinderella bonus points (these points are put in parentheses in the points earned column).

| **Team** | **Algorithm Rating** | **Tournament Games Won** | **Stats Showdown Points Earned** | **Conference** |
| --- | --- | --- | --- | --- |
| Liberty\* | 10 | 0 | 0 | Midwest |
| Baylor | 9 | 6 | 54 | South |
| Drake\* | 8 | 0 | 0 | West |
| Illinois | 7 | 1 | 7 | Midwest |
| Loyola Chicago | 6 | 2 | 12 | Midwest |
| Iowa | 5 | 1 | 5 | West |
| Gonzaga | 4 | 5 | 20 | West |
| Michigan | 3 | 3 | 9 | East |
| BYU | 2 | 0 | 0 | East |
| Ohio St. | 1 | 0 | 0 | South |

**Total Points Earned = 107 Maximum Points Possible = 655**

The first thing to will notice is that the total points earned with the adjusted algorithm is slightly worse, BUT we were able to get the 2 Championship teams (winning 5 and 6 games in the tournament) in this set and we have a total of 18 games won across all 10 teams, whereas our first iteration we only had 15.

Let’s see what happens if we test the algorithm on the 2013 through 2019 tournaments, the algorithm will perform at the following levels:

**2013 = 77 points 2014 = 96 points 2015 = 98 points**

**2016 = 100 points 2017 = 101 points 2018 = 142 points 2019 = 151 points**

The average points earned for this algorithm over the five tournaments is 109. By going with the adjusted algorithm there is a noticeable decrease in average performance. Back to the drawing board!!

### 

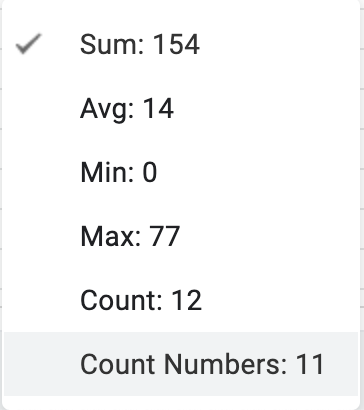
### Ideas for Analyzing Data in Spreadsheets

A major chunk of Data Analytics is scouring for information. The issue is not as difficult when you are dealing with a smaller set of data, but when you throw multiple spreadsheets with hundreds of data points into the mix, you need a more effective way of looking for data.

There are HUNDREDS of ways that you can analyze data within a singular data set. These are some examples, but you can search the internet to find even more ways.

*If you want to analyze the data yourself:*

* Using the filter function (and filter by) in the spreadsheet
* Sorting by columns
* Conditional formatting (especially using a color scale)
* Using the quick functions to determine sum, average, minimums and maximums



*If you want to use spreadsheet functions to help you analyze the data:*

* VLOOKUP function
* ABS function
* INDEX AND MATCH function
* Writing your own Macros

### Ideas for Developing Better Algorithms

This is just a sampling of things you can think about and types of algorithms you can use in your submission. Feel free to use any, all, or none of these ideas when developing your algorithm!

* **Look for patterns**
* **Iterate and fine-tune your algorithms**
* **Ask the right questions and then test your theories**
* **Use weighted values** (where certain data points are worth more or less than others)
* **Scale individual data sets to the interval [0,1]** (this allows you to more effectively compare apples, like percentages, to oranges, like counts) – this is a z-score in statistics
* **Look for correlations**
* **Develop ratings** (we did this in the 2nd iteration of our example algorithm when we ranked all teams from 68-1 for 2 data points)
* **Take advantage of the Cinderella Bonuses** (in the scoring section of this document, it shows that 2/3 of the total maximum possible points come from Cinderella points. While it is highly unlikely that situation would ever occur in real life, picking 1-3 well performing Cinderella teams in your algorithm can really help increase the overall scoring for your algorithm.)