How to Randomly Shuffle Rows in Excel for Better Data Analysis

Authored by

stats writer

February 14, 2026

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2026). *How to Randomly Shuffle Rows in Excel for Better Data Analysis*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=130581

The Fundamentals of Data Randomization in Microsoft Excel

In the expansive realm of **data analysis**, the ability to manipulate and reorganize information is a cornerstone of effective research and reporting. One specific operation that frequently arises is the need to **shuffle rows** in **Microsoft Excel**. Shuffling, or the process of rearranging the order of records randomly, is not merely a cosmetic change but a critical procedure used to ensure that the sequence of data does not introduce any hidden biases into your conclusions. By utilizing the robust features of a **spreadsheet**, users can transform a static list into a randomized dataset, which is particularly beneficial when preparing for **statistical sampling**, blinded studies, or even simple tasks like choosing a random winner from a list of contestants.

The core objective of shuffling is to eliminate any pre-existing patterns that might exist within the dataset. For instance, if a list of employees is sorted by hire date or alphabetical order, any subset taken from the top of that list would naturally be biased toward longer-tenured staff or specific surnames. By employing a **randomization** technique, every row is given an equal opportunity to appear in any position, thereby upholding the principles of **data integrity**. In **Excel**, this is traditionally achieved through the clever use of a **helper column** paired with a built-in function that generates volatile random numbers, allowing the software's internal **sorting algorithm** to do the heavy lifting.

Shuffling rows in **Excel** refers to rearranging the order of rows in a spreadsheet randomly. This feature is useful for data analysis and organization, as it allows for a different perspective on the data. To shuffle rows in **Excel**, one can use the **Sort** function or the **RAND** function in the **formula bar**. The **Sort** function allows for customizing the sorting criteria, while the **RAND** function generates a random number for each row, which can then be used to sort the rows in a random order. Both methods offer a quick and easy way to shuffle rows in **Excel**, providing a more efficient and effective approach to managing data.

Often you may want to randomly shuffle rows in **Excel**:

ARABPSYCHOLOGY.COM

	A	B	C	D	E	F	G	H
1	Original Rows					Shuffled Rows		
2	**Team**	**Points**	**Assists**			**Team**	**Points**	**Assists**
3	Mavs	22	4			Thunder	35	3
4	Spurs	19	9			Kings	15	8
5	Rockets	15	3			Jazz	33	2
6	Kings	15	8			Warriors	29	12
7	Warriors	29	12			Spurs	19	9
8	Nets	24	10			Mavs	22	4
9	Lakers	40	8			Rockets	15	3
10	Thunder	35	3			Nets	24	10
11	Blazers	23	6			Lakers	40	8
12	Jazz	33	2			Blazers	23	6
13								
14								
15								
16								

ARABPSYCHOLOGY.COM

Fortunately this is easy to do by using the **RAND()** function to create a **helper column** with random numbers for each row and then by sorting on the helper column. The following step-by-step example shows how to shuffle rows in **Excel** in practice.

Why Shuffling Rows is Essential for Rigorous Data Analysis

In professional environments, **data scientists** and researchers often require randomized datasets to validate their **hypotheses**. When data is collected, it often follows a chronological or categorical sequence that can skew results if not properly addressed. For example, in **A/B testing**, it is vital to assign participants to groups randomly to ensure that the groups are comparable. Shuffling the rows of your primary **database** or **Excel** sheet before group assignment is a standard practice to mitigate the risk of **confounding variables** affecting the outcome of the experiment.

Furthermore, shuffling serves as an excellent tool for **data visualization**. Sometimes, viewing data in its original sorted state can make it difficult to spot outliers or unique trends that only become apparent when the records are mixed. By viewing a randomized version of a **spreadsheet**, an analyst might notice clusters or anomalies that were previously obscured by a rigid alphabetical or numerical sort. This fresh perspective is often the catalyst for new insights in **exploratory data analysis**, providing a more comprehensive understanding of the underlying information.

Finally, the process of shuffling in **Excel** is highly valued for its simplicity and accessibility. Unlike complex statistical software that may require extensive coding knowledge, **Excel** provides a user-

friendly interface to achieve high-quality randomization. Using the **RAND function** allows even novice users to perform sophisticated data manipulation tasks. This accessibility ensures that **data-driven decision-making** is not limited to those with advanced programming skills but is available to anyone capable of navigating the **Excel ribbon** and the **Data tab**.

Step 1: Establishing and Preparing Your Initial Dataset

The first prerequisite for successfully shuffling rows is to have a well-structured and clean dataset. Before applying any **formulas**, ensure that your data is organized in a **tabular format**, where each row represents a unique record and each column represents a specific attribute. It is also advisable to ensure there are no completely empty rows or columns within your data range, as these can sometimes disrupt the **Sort** functionality, leading to fragmented results where only a portion of the data is actually moved.

First, let's create the following dataset that contains information about various basketball players, including their names, team affiliations, and points scored:

	A	B	C	D	E	F
1	**Team**	**Points**	**Assists**			
2	Mavs	22	4			
3	Spurs	19	9			
4	Rockets	15	3			
5	Kings	15	8			
6	Warriors	29	12			
7	Nets	24	10			
8	Lakers	40	8			
9	Thunder	35	3			
10	Blazers	23	6			
11	Jazz	33	2			
12						
13						
14						
15						
16						

ARABPSYCHOLOGY.COM

When preparing the dataset, it is crucial to consider whether you want to include the **header row** in your shuffle. Most users prefer to keep the headers at the top of the sheet to maintain context. **Excel** has an "My data has headers" option in the **Sort dialog box** that accounts for this, but manually checking your data range (e.g., A2:C11 instead of A1:C11) is a reliable way to ensure

your labels stay exactly where they belong while the underlying data is randomized.

Additionally, if your dataset contains **merged cells**, you should unmerge them before attempting to shuffle. **Sorting** operations in **Excel** generally require all cells within the sort range to be of uniform size. If the software encounters merged cells, it will often produce an error message or fail to execute the sort correctly. Cleaning these elements beforehand guarantees a smooth transition through the subsequent steps of the randomization process and maintains the **data quality** of your final output.

Step 2: Utilizing the RAND Function to Create a Helper Column

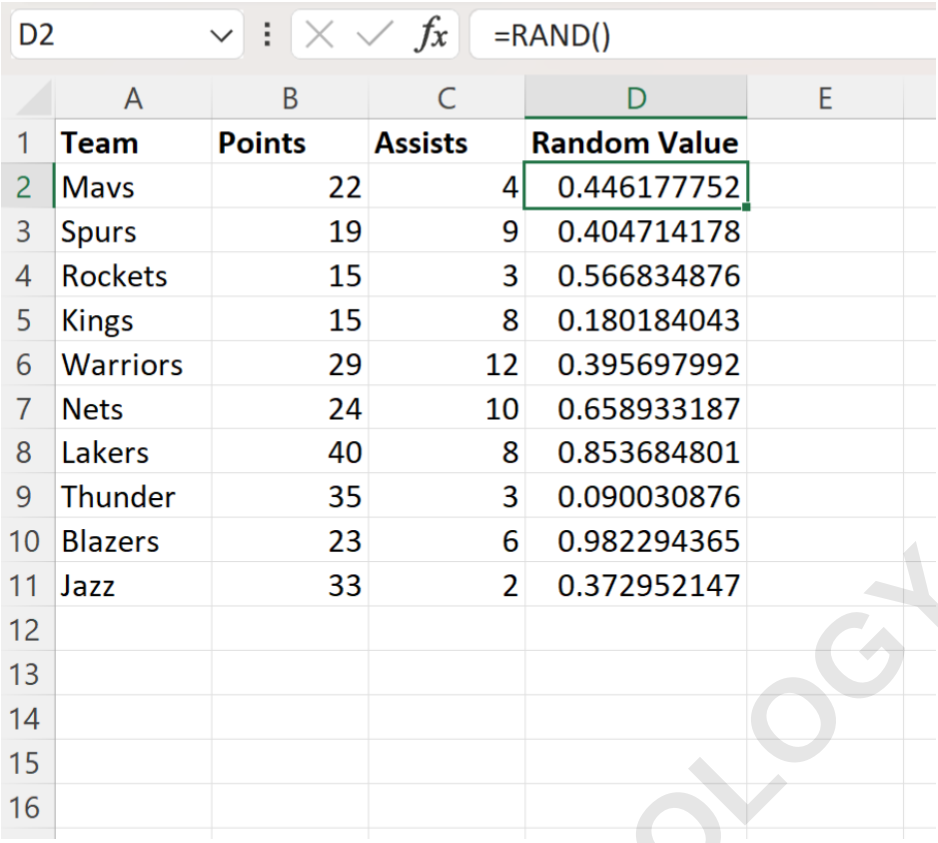
The core mechanism for shuffling in **Excel** involves creating a **helper column**. This column acts as a temporary anchor that provides a unique, random value for every single row in your dataset. Because **Excel** does not have a single "Shuffle" button, we use the **RAND function** to assign a decimal value between 0 and 1 to each record. Since these numbers are generated by a **pseudorandom number generator**, they are effectively unique and lack any logical connection to the existing data in the row.

Next, type the following formula into cell **D2** to generate a random value between 0 and 1:

=RAND()

We can then click and drag this **fill handle** down to each remaining cell in column D. As you apply the formula, you will notice that **Excel** immediately populates the cells with various decimal numbers. These numbers are "volatile," meaning they will recalculate every time you make a change to the **spreadsheet** or press the F9 key. This volatility is exactly what we need for a one-time shuffle, as it ensures the values are fresh and unbiased before the **sorting** operation begins.

ARABPSYCHOLOGY.COM



	A	B	C	D	E
1	**Team**	**Points**	**Assists**	**Random Value**	
2	Mavs	22	4	0.446177752	
3	Spurs	19	9	0.404714178	
4	Rockets	15	3	0.566834876	
5	Kings	15	8	0.180184043	
6	Warriors	29	12	0.395697992	
7	Nets	24	10	0.658933187	
8	Lakers	40	8	0.853684801	
9	Thunder	35	3	0.090030876	
10	Blazers	23	6	0.982294365	
11	Jazz	33	2	0.372952147	
12					
13					
14					
15					
16					

ARABPSYCHOLOGY.COM

It is worth noting that while **RAND()** is the most common choice, some users might prefer the **RANDBETWEEN** function if they require specific integer ranges. However, for the purpose of shuffling, **RAND()** is superior because the likelihood of generating duplicate values (known as "collisions") is extremely low due to the high number of decimal places. This ensures that every row has a distinct value, which allows the **sorting algorithm** to arrange them in a truly random sequence without having to decide how to handle "tied" values.

Understanding the Mechanics of Random Number Generation

To appreciate how the shuffle works, it is helpful to understand the **RAND function**. In **computer science**, generating truly random numbers is a complex task. **Excel** uses an algorithm to produce these values, which are technically **pseudorandom**. For standard business and data analysis purposes, these values are more than sufficient to achieve a high degree of **entropy**. When you enter **=RAND()**, you are essentially asking **Excel** to pull a value from its internal sequence that mimics randomness.

One important characteristic of the **RAND function** is its **volatility**. Every time a calculation occurs in the workbook, the **RAND** values change. If you want to "lock" the random order after shuffling, you would need to use the **Paste Values** feature to convert the formulas into static numbers. However, for a simple shuffle, this is not strictly necessary as the act of sorting will finalize the

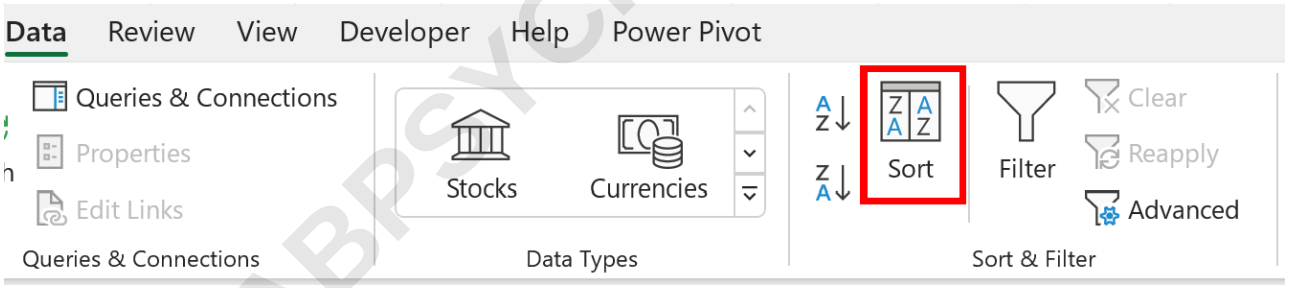
order based on the values present at the exact moment the **Sort** command is executed.

Using a **helper column** is a non-destructive way to manipulate your data. By adding a new column rather than modifying existing ones, you preserve the **metadata** and the original values of your basketball players' stats. This approach is a best practice in **data management**, as it allows you to easily audit your work or revert the changes if something goes wrong during the sorting process. Once the rows are in their new random positions, the helper column has fulfilled its purpose and can be removed without affecting the randomized order of the other columns.

Step 3: Executing the Shuffle via Sorting Tools

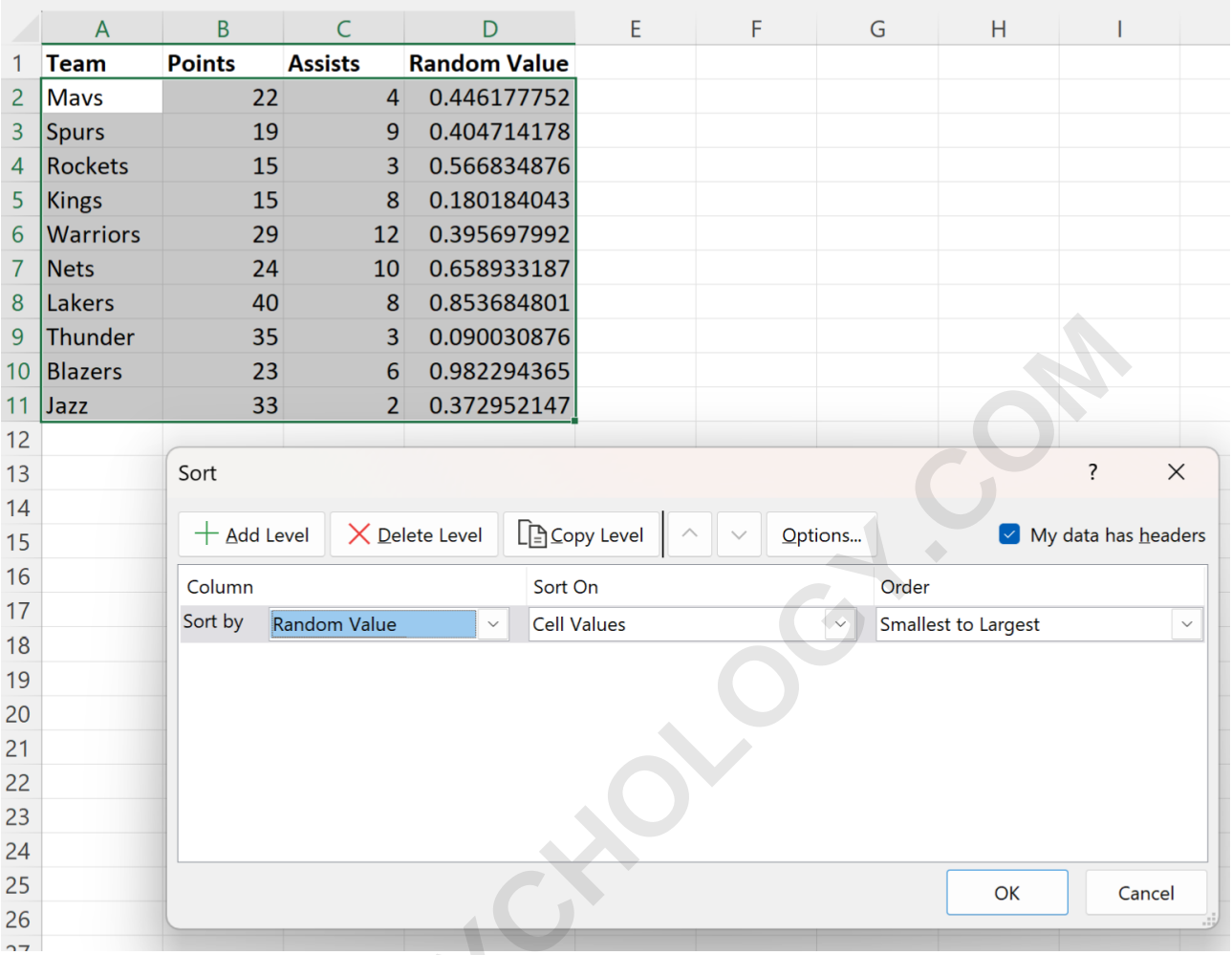
With the random values now in place, the next step is to perform the actual reorganization of the rows. This is done through the **Sort** dialog in the **Data tab** of the **Excel ribbon**. By sorting based on the **helper column**, you are telling **Excel** to ignore the names, teams, or points and instead focus entirely on the random decimals. Because those decimals were distributed without any pattern, the resulting order of the basketball players will be completely randomized.

We can now "shuffle" the rows in the dataset by sorting the rows based on the values in the **Random Value** column. To do so, highlight the cell range **A2:D11**, then click the **Sort** icon within the **Sort & Filter** group in the **Data** tab along the top ribbon:



In the new **Sort** window that appears, choose **Random Value** in the **Sort by** dropdown menu. You can choose to sort from "Smallest to Largest" or "Largest to Smallest"; since the numbers are random, the direction of the sort does not matter. Both options will result in a successfully shuffled list of data. This step effectively uses the **sorting algorithm** to re-index your data according to the random numbers generated in Step 2.

ARABPSYCHOLOGY.COM



Finalizing the Randomization and Cleaning the Workbook

Once you click **OK**, the rows will be sorted from smallest to largest based on the values in the **Random Value** column. You will see the names of the basketball players move to new positions instantly. The **user interface** of **Excel** handles the relocation of the entire row, ensuring that "Player A" always stays with their respective "Team" and "Points," maintaining the **relational integrity** of the dataset. This is why it is vital to select the entire range of data before sorting, rather than just the helper column.

ARABPSYCHOLOGY.COM

	A	B	C	D	E
1	**Team**	**Points**	**Assists**	**Random Value**	
2	Thunder	35	3	0.243777292	
3	Kings	15	8	0.281144584	
4	Jazz	33	2	0.404608717	
5	Warriors	29	12	0.422129311	
6	Spurs	19	9	0.22088698	
7	Mavs	22	4	0.63767251	
8	Rockets	15	3	0.369562397	
9	Nets	24	10	0.584550037	
10	Lakers	40	8	0.411538687	
11	Blazers	23	6	0.994781735	
12					
13					
14					
15					

ARABPSYCHOLOGY.COM

The original rows in the dataset have now been "shuffled" in a random order. At this point, the **Random Value** column has served its purpose. To clean up your **spreadsheet** and make it ready for presentation or further analysis, you can simply right-click the header of Column D and select **Delete**. Your data will remain in its new, shuffled order. It is important to remember that because the **RAND function** is volatile, the values in the helper column might have changed one last time during the sort, but the final arrangement of your data rows is now fixed.

Note: Feel free to delete the **Random Value** column after you've sorted the rows since you no longer need it. This keeps your **worksheet** tidy and focused on the relevant **business intelligence** or research data. If you ever need to shuffle the data again, you can simply repeat the process by creating a new helper column and reapplying the **RAND function**.

Modern Alternatives: Using SORTBY and RANDARRAY in Excel 365

For users who are utilizing **Microsoft 365** or **Excel 2021** and later, there is a more modern and dynamic way to shuffle rows without the need for manual sorting or helper columns. The introduction of **dynamic arrays** brought about the **SORTBY** and **RANDARRAY** functions. These functions allow you to create a shuffled version of your data that updates automatically. This is particularly useful for **dashboards** or reports where you want a fresh random sample every time the file is opened.

The **SORTBY** function works by taking an array (your data) and sorting it based on the values in

another array. By using **RANDARRAY** as the second argument, you generate a temporary list of random numbers in memory that **Excel** uses to determine the new order. The syntax looks something like this: `=SORTBY(A2:C11, RANDARRAY(ROWS(A2:C11)))`. This **formula** is extremely powerful because it leaves your original dataset untouched while displaying the shuffled results in a new location.

This method is often preferred by **power users** who want to maintain a "source of truth" in one part of their **workbook** while displaying various randomized views elsewhere. It also avoids the manual steps of clicking through the **Data tab** and deleting helper columns. However, the traditional method remains the most universal, as it works across all versions of **Excel**, including legacy versions often found in corporate environments, ensuring that your skills are transferable regardless of the specific **software version** you are using.

Best Practices for Maintaining Data Integrity During Shuffles

When performing any major operation like a shuffle, **data backups** are essential. Before you begin the process, it is wise to save a copy of your **workbook** or duplicate the **worksheet**. This ensures that if you accidentally sort only one column (thereby mismatching your data) or if the **sorting algorithm** behaves unexpectedly, you have a clean version to return to. In **data management**, preserving the original state of raw data is a fundamental rule that prevents permanent errors.

Another best practice is to verify the **cell references** in any existing formulas within your dataset. If your rows contain formulas that point to other rows (such as a running total), shuffling the rows may cause those formulas to break or produce incorrect results. Before shuffling, consider converting those formulas to static values using the **Paste Special > Values** command. This "freezes" the data in place, ensuring that the **points** or other metrics remain accurate for each basketball player regardless of their new position in the list.

Lastly, always double-check the **Sort range**. One of the most common mistakes in **Excel** is failing to include all columns in the sort. If you only select the "Player Name" and "Random Value" columns but forget the "Team" and "Points" columns, the players will move but their stats will stay in the original order. This results in **corrupted data**. By highlighting the entire table or using the **Ctrl+A** shortcut to select the current region, you ensure that **Excel** treats each row as a single, unbreakable unit during the shuffle.

Advanced Applications of Randomized Data

Beyond simple organization, the ability to shuffle rows opens the door to more advanced **statistical techniques** within **Excel**. For instance, once your data is shuffled, you can easily perform **systematic sampling** by selecting every 10th row. Since the list is already in a random order, this selection method provides a **random sample** that is statistically valid for making

ARABPSYCHOLOGY.COM

inferences about a larger population. This is a common requirement in **quality control** and auditing processes.

In addition to sampling, randomized rows are useful for **machine learning** preparation. When training a model, it is standard practice to shuffle the dataset before splitting it into **training and testing sets**. This ensures that the model does not learn any accidental patterns based on the order of data entry. **Excel** serves as an excellent environment for this type of initial **data preprocessing**, allowing analysts to prepare their data before exporting it to more specialized programming languages like **Python** or **R**.

Ultimately, mastering the shuffle technique enhances your overall **Excel proficiency**. It demonstrates an understanding of how to combine **logical functions** with built-in **data tools** to achieve complex results. Whether you are managing a sports league, conducting academic research, or analyzing business trends, the ability to effectively randomize information is a versatile skill that contributes to more robust and reliable **data analysis** outcomes.

The following tutorials explain how to perform other common operations in **Excel**:

How to Select a Random Sample in Excel
How to Use the RANDBETWEEN Function
Advanced Sorting Techniques for Large Datasets

ARABPSYCHOLOGY.COM