How to Normalize Data: Scaling Values Between 0 and 100

Authored by

Mohammed looti

November 6, 2025

Data preprocessing stands as a critical step in nearly all quantitative fields, including statistical analysis and machine learning model development. Among the various techniques used to condition raw data, normalization is perhaps the most fundamental, serving to scale numerical features to a standardized range. This article provides an in-depth focus on a specific, highly practical application of this technique: transforming raw values into a scale ranging precisely between 0 and 100. This 0-100 scale is frequently employed when constructing composite indices, standardized metrics, or easily interpretable scoring systems, as it mirrors the familiar concept of percentages.

The consistent scaling of data is essential because it ensures that all variables contribute fairly and equally to the subsequent analysis. Without proper scaling, features that naturally possess large numerical ranges--such as income measured in thousands--can disproportionately dominate the calculations of distance-based algorithms, effectively masking the true impact of features with smaller ranges, such as age or rating scores. The approach discussed here, often formally referred to as Min-Max normalization, is favored for its simplicity and its ability to map the entire distribution of the original data onto a universally understood percentage scale, ensuring immediate interpretability of the resulting scores.

The Min-Max Scaling Formula (0 to 100)

To successfully transform data points within a dataset so that they fall within the 0 to 100 boundaries, we employ a specific adaptation of the standard Min-Max scaling formula. The core mechanism of this process relies on identifying the absolute minimum and maximum values within the original range; the formula then maps the original minimum to 0 and the original maximum to 100, scaling all intermediate values proportionally based on their position relative to those range boundaries. This process ensures that the inherent relationships and distribution shape of the data are preserved, though the scale is completely redefined.

The mathematical foundation for scaling data to a target range of 0 to 100 is defined precisely by the following equation, which forms the basis for all calculations throughout this process:

zi = (xi - min(x)) / (max(x) - min(x)) * 100

For accurate implementation and true mastery of this technique, a solid understanding of each variable within the formula is critical. The equation operates by first calculating the relative position of the data point within the original range, dividing the difference between the point and the minimum by the total range span, and finally multiplying this resulting ratio by 100 to achieve the desired percentage scale.

zi: Represents the ith normalized score, which will reside within the 0-100 range.
xi: Represents the ith original raw value retrieved directly from the input dataset.

min(x): The absolute **minimum value** observed across the entirety of the original dataset.
max(x): The absolute **maximum value** observed across the entirety of the original dataset.

Step-by-Step Example: Scaling Data to 0-100

To solidify the theoretical understanding of the Min-Max formula, let us apply this normalization technique to a practical, simple dataset. Imagine we are working with a list of raw scores where the values vary considerably, and our objective is to convert these raw figures into a standardized percentage scale for comparative analysis and reporting purposes. This transformation will allow us to immediately interpret whether a score is closer to the lowest performance (0) or the highest performance (100).

Consider the following raw dataset, which contains five distinct measurements:

Data Values
12
19
21
23
25
35
47
48
59
65
66
67
68

The first essential step is to inspect the values to establish the definitive boundaries of the original data spread. In this specific dataset, we can clearly identify that the **minimum value** is 12, and the **maximum value** is 68. These two anchor points (min=12, max=68) are constants that define the total range, which is calculated as 68 - 12 = 56. This range (56) will form the static denominator for every single calculation we perform during this normalization process.

We will now meticulously demonstrate how the transformation process works for the first three data points using the 0-100 formula, ensuring that the consistency of the boundaries (min=12, max=68) is maintained throughout:

Normalizing the Minimum Value (12): When we apply the formula to the smallest original value, **12**, we are confirming the lower boundary of our new range:

zi = (xi - min(x)) / (max(x) - min(x)) * 100 = (12 - 12) / (68 - 12) * 100 = **0**

This result confirms the integrity of the method: the minimum value in the original scale maps perfectly to 0 in the normalized percentage scale.

Normalizing the Second Value (19): Applying the exact same established logic and range boundaries to the second value, **19**, allows us to determine its proportional position:

zi = (xi - min(x)) / (max(x) - min(x)) * 100 = (19 - 12) / (68 - 12) * 100 = **12.5**

Normalizing the Third Value (21): Finally, for the third value of **21**, the calculation shows a slightly higher proportional score:

zi = (xi - min(x)) / (max(x) - min(x)) * 100 = (21 - 12) / (68 - 12) * 100 = **16.07**

By systematically applying this formula to every single value in the original dataset, the entire distribution is transformed into a set of new, standardized scores that strictly adhere to the range between 0 and 100. This comprehensive transformation is displayed in the resulting table below, providing a clear visual representation of the scaled data:

Data Values	Normalized
12	0
19	12.5
21	16.07
23	19.64
25	23.21
35	41.07
47	62.5
48	64.29
59	83.93
65	94.64
66	96.43
67	98.21
68	100

Generalizing the Scale: Normalizing to Any Range (0 to Q)

While scaling data to the 0-100 range is exceptionally useful for percentage-based scoring, the Min-Max formula offers remarkable versatility that extends far beyond this standard. It can be readily adapted to scale data between 0 and any arbitrary positive upper boundary, which we shall designate as **Q**. This flexibility is invaluable when developing highly specialized indices, proprietary metrics, or when working within systems that require a normalized range other than the typical

percentage scale.

The generalized formula is a straightforward conceptual extension of the previous equation. It maintains the core structure of calculating the data point's relative position within the original range, but replaces the constant multiplier of 100 with the desired maximum value, **Q**:

zi = (xi - min(x)) / (max(x) - min(x)) * Q

In this generalized context, **Q** represents the absolute maximum target value for your normalized data. In our initial example, we fixed Q at 100. However, the same original range of data values could easily be normalized to fall between 0 and 1,000 simply by selecting Q to be 1,000. This demonstrates that the scaling factor dictates the final spread, while the relative distance between data points remains constant.

Continuing with our previous dataset (where min=12 and max=68), let's observe how the numerical results change when the target maximum **Q** is set to 1,000, effectively stretching the scale by a factor of ten compared to the 0-100 normalization:

To normalize the original minimum value of **12**, we apply the updated formula:

zi = (xi - min(x)) / (max(x) - min(x)) * 1,000 = (12 - 12) / (68 - 12) * 1,000 = **0**

To normalize the second value of **19**, we use the new scaling factor:

zi = (xi - min(x)) / (max(x) - min(x)) * 1,000 = (19 - 12) / (68 - 12) * 1,000 = **125**

To normalize the third value of **21**, the calculation yields a substantially larger score than before:

zi = (xi - min(x)) / (max(x) - min(x)) * 1,000 = (21 - 12) / (68 - 12) * 1,000 = **160.7**

These calculations clearly illustrate that the fundamental proportional relationship among the data points is preserved, but the entire scale is stretched due to the selection of **Q**. Applying this across the entire dataset produces a new range of values strictly bounded between 0 and 1,000, as demonstrated below:

Data Values	Normalized
12	0
19	125
21	160.7
23	196.4
25	232.1
35	410.7
47	625
48	642.9
59	839.3
65	946.4
66	964.3
67	982.1
68	1000

Context and Importance in Data Science

Grasping the underlying rationale for **when** and **why** data normalization is necessary is arguably as important as understanding the formula itself. Normalization is typically mandatory when performing statistical or machine learning analysis that involves combining multiple variables measured across fundamentally different scales, units, or orders of magnitude. For instance, comparing a feature measured in meters (e.g., height) with a feature measured in square kilometers (e.g., area) necessitates a common scale to avoid distorting the analytical outcome.

The core objective of normalization is the establishment of parity among all variables. Consider a scenario where an analysis includes 'income' (measured potentially in hundreds of thousands) alongside 'age' (measured only in decades). Algorithms that rely on calculating geometric distance between data points, such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), or clustering algorithms, will inherently allow the variable with the larger magnitude (income) to exert a vastly disproportionate influence on the final model results. This occurs because the distance metric is dominated by the scale of the largest feature.

By scaling all variables, typically to a common range such as 0 to 1 or 0 to 100, we effectively eliminate this measurement bias. This crucial step guarantees that the analytical outcomes accurately reflect the true, underlying statistical relationships within the data, rather than being mere artifacts resulting from arbitrary choices of measurement units. Normalization ensures that the model learns from the importance of the features, not the size of their values.

Min-Max vs. Standardization (Z-Score)

It is important to recognize that normalization is a broad term encompassing several distinct

techniques for feature scaling. While this tutorial focuses heavily on Min-Max Normalization, the two most ubiquitous methods used in data preprocessing are Min-Max scaling and Mean Normalization, often known interchangeably as Standardization or Z-Score scaling. The selection between these methods is a critical decision, heavily dependent on the specific requirements of the analytical model, the presence of outliers, and the underlying distribution of the data.

1. Min-Max Normalization (Scaling to a Fixed Range)

As extensively detailed in this article, Min-Max scaling is the optimal technique when a rigid, fixed boundary is a prerequisite for the analysis, such as when generating bounded performance scores, visualization scales, or indices that must range exactly from 0 to 100. Although it preserves the original distribution shape, it is sensitive to extreme outliers, as a single large outlier can significantly compress the rest of the data points into a very small portion of the new scale.

Objective: To convert every data value into a scaled value that is strictly contained within a user-defined range (e.g., 0 and 100).
Formula: New value = (value - min) / (max - min) * Range Max

2. Mean Normalization (Standardization / Z-Score)

Conversely, Mean Normalization is frequently favored, particularly in robust machine learning applications, when the primary goal is to center the data distribution around zero, irrespective of the original minimum and maximum values. This process transforms the data into a distribution with a mean of 0 and a standard deviation of 1. Standardization is particularly beneficial for algorithms that assume a Gaussian (normal) distribution or that perform better when features are centered, and it tends to be less affected by extreme outliers than Min-Max scaling.

Objective: To scale values such that the **mean** of the entire resulting dataset is 0 and the **standard deviation** is 1.
Formula: New value = (value - mean) / (standard deviation)

While both Min-Max normalization and Standardization are essential tools in the data preprocessing toolkit, Min-Max normalization remains the definitive choice whenever a specific, absolute, and bounded output range--such as 0 to 100--is mandated for the resulting scores or metrics.

Additional Resources for Data Scaling

For readers who wish to further explore the practical implementation of data scaling techniques across various environments and programming languages, the following resources provide

additional guidance and specific technical applications:

How to Normalize Data Between 0 and 1
How to Normalize Data in Excel
How to Normalize Data in R
How to Normalize Columns in Python