

8-11-2024

Training Day – 37

Topic:* Standardizing Data

- Applied transformations to ensure consistency in data formatting.
- Example: Converted all text columns to uppercase using `.str.upper()`.

Data standardization is an important technique that is mostly performed as a pre-processing step before inputting data into many machine learning models, to standardize the range of features of an input data set.

Standardizing Data in Excel

Excel STANDARDIZE is available under Excel Statistical Functions. It returns a normalized value, which is also called Z-score.

The mean and standard deviation are the basis of the z-score. The z-score (or standard score) is a method to standardize scores across the same scale. It divides a score's deviation by the standard deviation in a data set. The resulting score is the standard deviation of a data point from the mean.

Zero is the average of all z-scores for a dataset. A negative z score indicates that the value is lower than the mean. A positive z score indicates that the value is higher than the mean.

Z-Score Formula = STANDARDIZE(x, mean, standard_dev)

Here: X= data value that you need to normalize.

Mean= Distribution arithmetic mean

Standard_dev= Distribution standard deviation.