## Feature Scaling Technique for Normalization: StandardScaler

The technique of normalization used in the normalization of this analysis was StandardScaler which is a method of the scikit-learn library. The approach standardizes variables by finding the mean and dividing by the standard deviation that is set to one. It is especially appropriate with algorithms such as K-Means, where the scale of input features is an issue because they use Euclidean distance to measure similarity.

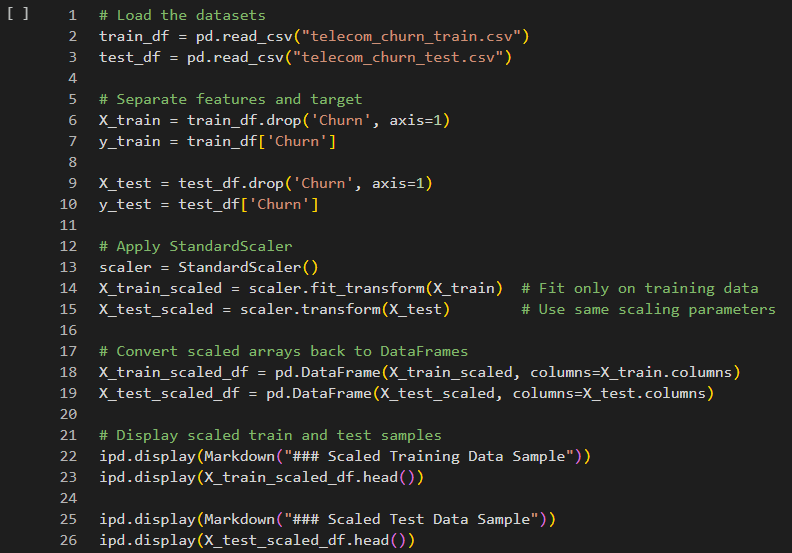
Each feature 𝑥 is transformed using the formula:

​

where 𝜇 is the mean and 𝜎 is the standard deviation of the feature of training set.

The scaler was only applied to the training data and test data was not exposed to the transformation. These parameters of transformation (mean and standard deviation) were then used on the test set. This way prevents data leakage and consistency in scaling between the two datasets. Standardization contributed to the fact that all features were put on the same scale, which allowed more equal clustering without the predominance of one variable.

## Code Snippet



## Output of Scaling

